

SRI International

AD-A253 908**DTIC**
ELECTE
AUG 4 1992
S C D

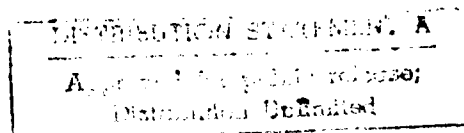
Final Report • February 1992

MECHANICAL PROPERTIES OF SEMICONDUCTORS AND THEIR ALLOYS

A. Sher, Associate Director
M.A. Berding, Research Physicist
A.T. Paxton, International Fellow
S. Krishnamurthy, Research Physicist
Physical Electronics Laboratory

A.-B. Chen
Auburn University
Auburn, Alabama

SRI Project 6682



Prepared for:

Office of Scientific Research
United States Air Force
Bolling Air Force Base, Building 410
Washington, DC 20332-6448

Attn: Dr. Gernot Pomrenke

Contract F49620-88-K-0009

92-20462**92 7 28 105**

MECHANICAL PROPERTIES OF SEMICONDUCTORS AND THEIR ALLOYS

A. Sher, Associate Director
M.A. Berding, Research Physicist
A.T. Paxton, International Fellow
S. Krishnamurthy, Research Physicist
Physical Electronics Laboratory

A.-B. Chen
Auburn University
Auburn, Alabama

SRI Project 6682

Prepared for:

Office of Scientific Research
United States Air Force
Bolling Air Force Base, Building 410
Washington, DC 20332-6448

Attn: Dr. Gernot Pomrenke

Contract F49620-88-K-0009

Approved:

Ivor Brodie, Director
Physical Electronics Laboratory

Donald L. Nielson, Vice President
Computing and Engineering Sciences Division

DTIC QUALITY INSPECTED 5

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188						
REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS							
SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT							
DECLASSIFICATION / DOWNGRADING SCHEDULE			Unlimited							
PERFORMING ORGANIZATION REPORT NUMBER(S) 6682 Final Report			5. MONITORING ORGANIZATION REPORT NUMBER(S)							
NAME OF PERFORMING ORGANIZATION UI International		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NE							
ADDRESS (City, State, and ZIP Code) 3 Ravenswood Avenue Emery Park, California 94025		7b. ADDRESS (City, State, and ZIP Code) Bldg 410 Bolling AFB DC 20332								
NAME OF FUNDING / SPONSORING ORGANIZATION AF Office of Scientific Research		8b. OFFICE SYMBOL (If applicable) NE	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-88-K-0009							
ADDRESS (City, State, and ZIP Code) Bolling Air Force Base Washington, DC 20332-6448		10. SOURCE OF FUNDING NUMBERS								
		<table border="1"><tr><td>PROGRAM ELEMENT NO. 61102F</td><td>PROJECT NO. 2306 6682</td><td>TASK NO. B1</td><td colspan="2">WORK UNIT ACCESSION NO.</td></tr></table>				PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2306 6682	TASK NO. B1	WORK UNIT ACCESSION NO.	
PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2306 6682	TASK NO. B1	WORK UNIT ACCESSION NO.							
TITLE (Include Security Classification) Mechanical Properties of Semiconductors and Their Alloys										
PERSONAL AUTHOR(S) A. Sher, M.A. Berding, A.T. Paxton, S. Krishnamurthy, A.-B. Chen										
TYPE OF REPORT Final		13b. TIME COVERED FROM 9-1-88 TO 1-13-92		14. DATE OF REPORT (Year, Month, Day) 1992 February						
				15. PAGE COUNT 338						
SUPPLEMENTARY NOTATION										
COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)							
FIELD	GROUP	SUB-GROUP	Mechanical properties, semiconductors, first principles, FP-LMTO, LCAO, tight binding							
ABSTRACT (Continue on reverse if necessary and identify by block number)										
<p>A wide range of subjects have been treated in this contract. We have devoted time to the development and applications of two first-principles computational methods: one, the full-potential linear muffin tin orbital (FP-LMTO) method is somewhat mature and highly accurate, while the other, linear combination of atomic orbitals (LCAO), is less accurate but more flexible and is easily incorporated into the other calculations we have in place, e.g., surface Green's function methods and CPA. Tight binding has also been used.</p> <p>These methods have been applied to solve a host of mechanical-property problems including elastic constants, cleavage energies, substitution energies, interactions between surface atoms relating to their surface order-disorder state and growth theory, surface segregation, k order-disorder theory and phase stability, the effect of dislocations on electronic transport and electro-optic properties of semiconductors, the Ni-Al intermetallic phase diagram, planar fault energies in L12 alloys, high-performance structural metal alloy design, and a contribution to understanding the Jones theory of metal alloying. Many of these subjects have been brought to publishable conclusions. Whenever possible, we have presented our detailed results in the form of preprints and reprints, with only brief summaries of the work given elsewhere. In instances where the research is incomplete, we have given somewhat longer expositions.</p>										
DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified							
NAME OF RESPONSIBLE INDIVIDUAL Dr. Gernot S. Pomrenke			22b. TELEPHONE (Include Area Code) (202) 767-4931		22c. OFFICE SYMBOL NE					

CONTENTS

1.	INTRODUCTION	1
2.	SEMICONDUCTORS.....	2
2.1	Development of <i>Ab Initio</i> Theories	2
2.1.1	KKR Theory	2
2.1.2	Linear Combination of Atomic Orbitals: Theory	2
2.1.3	Linear Muffin Tin Orbitals: Theory	7
2.2	Theory of Elastic Constants.....	8
2.3	Surfaces, Interfaces, and Superlattices.....	10
2.3.1	Difference-Equation Approach	10
2.3.2	Cleavage Energy	11
2.3.3	Energetics of Atoms on Surfaces	11
2.3.4	Surface Segregation in Alloys	12
2.4	Order-Disorder Theory	12
2.4.1	Formation Energies and Phase Stability	12
2.4.2	Critical Temperature of Order-Disorder Transitions	13
2.4.3	Cluster Energies	13
2.5	Dislocations and Electro-optic Properties	13
3.	FIRST-PRINCIPLES APPROACH TO THE PLASTIC PROPERTIES OF HIGH-TEMPERATURE ALLOYS	15
3.1	Introduction.....	15
3.2	The Ni-Al Phase Diagram.....	15
3.3	Planar Fault Energies in L1 ₂ Alloys.....	15
3.4	Quantum Mechanical Alloy Design and the Transition-Metal Trialuminides	20
3.5	The Jones Theory of Alloying	20
	REFERENCES	21
	APPENDICES	
A	Full-Potential Korringa-Kohn-Rostoker Band Theory Applied to the Mathieu Potential.....	A-1
B	Defect Equilibrium in HgTe	B-1
C	Formation Energies, Bond Lengths, and Bulk Moduli of Ordered Semiconductor Alloys from Tight-Binding Calculations.....	C-1
D	Elastic Constants and Related Properties of Semiconductor Compounds and Their Alloys	D-1
E	Electronic Structures of HgTe and CdTe Surfaces and HgTe/CdTe Interfaces	E-1
F	Alloy Statistics and Phase Diagrams	F-1
G	HgCdTe Status Review with Emphasis on Correlations, Native Defects, and Diffusion	G-1
H	Jones' Theory of the Hume-Rothery Phases Revisited.....	H-1
I	First Principles Determination of the Ni-Al Phase Diagram	I-1
J	Al ₃ Ru—A Ductile Trialuminide?	J-1

FIGURES

1 Cohesive energies and shear elastic coefficients: comparison between theory and experiment	9
2 Calculation of shear modulus (circles) and bulk modulus (squares) for superlattices of CdZnTe along [110].....	10
3 (111) planar fault energies in three L1 ₂ alloys as a function of in-plane translation	18

TABLES

1 Structural Energy Differences in Pt ₃ Al and Cu ₃ Au	16
2 APB Energy in Pt ₃ Al as Function of Supercell Size	17
3 Characters of Planar Faults in the L1 ₂ Lattice	17
4 Planar Fault Energies $g(\text{J/m}^2)$ and Translation States in L1 ₂ Intermetallics.....	19

1. INTRODUCTION

A wide range of subjects have been treated in this contract. We have devoted time to the development and applications of two first-principles computational methods: one, the full-potential linear muffin tin orbital (FP-LMTO) method is somewhat mature and highly accurate, while the other, linear combination of atomic orbitals (LCAO), is less accurate but more flexible and is easily incorporated into the other calculations we have in place, e.g., surface Green's function methods and CPA. Tight binding has also been used.

These methods have been applied to solve a host of mechanical-property problems including elastic constants, cleavage energies, sublimation energies, interactions between surface atoms relating to their surface order-disorder state and growth theory, surface segregation, bulk order-disorder theory and phase stability, the effect of dislocations on electronic transport and electro-optic properties of semiconductors, the Ni-Al intermetallic phase diagram, planar fault energies in $L1_2$ alloys, high-performance structural metal alloy design, and a contribution to understanding the Jones theory of metal alloying. Many of these subjects have been brought to publishable conclusions. Whenever possible, we have presented our detailed results in the form of preprints and reprints, with only brief summaries of the work given here. In instances where the research is incomplete, we have given somewhat longer expositions.

2. SEMICONDUCTORS

2.1 DEVELOPMENT OF *AB INITIO* THEORIES

2.1.1 KKR Theory

Many of our previous calculations were based on the empirical tight-binding method (ETB). This method is most useful as an interpolation scheme in which the properties of reference materials are well known. The accuracy is expected to deteriorate if ETB is used to extrapolate too far from known systems. For example, ETB should be reliable for alloy systems because the end points of the alloys, namely the constituent compounds, are well described by ETB and the alloy structures are very similar to those of the hosts. However, we are not as confident regarding surface calculations because the broken bonds at the surface are very different from the tetrahedral bonds of the bulk. For surface problems, an *ab initio* approach is more suitable.

In addition to the full-potential LMTO method which has been established at SRI (see Section 2.1.3), efforts have been made to develop an *ab initio* LCAO scheme and a multiple-scattering KKR (MSKKR) Green's function theory (Yeh et al., 1990; Appendix A) using space-filling potentials. The reason for studying LCAO is that it can be transformed into the ETB form for which we have developed many useful techniques that can be employed directly, such as the CPA and the difference equation approach (see Section 2.3.1). Similarly, once we can deal with nonspherical potentials in Green's function formalism, we can extend the KKR-CPA and defect calculations using self-consistent potentials to semiconductors and alloys.

A numerical test of the MSKKR theory using the Mathieu potential shows excellent convergence for bands up to 1.5 Ry (Yeh et al., 1990). This method has been implemented in a total energy calculation for silicon and III-V compounds within the Harris functional approximation. Results from the first successful test of the LCAO on diatomic molecules have been obtained. The extension to bulk semiconductors using limited basis and the Harris approximation has also been obtained. To make these approaches applicable to surfaces, more work incorporating self-consistency with an expanded basis set is needed.

2.1.2 Linear Combination of Atomic Orbitals: Theory

The empirical band structures, even when they are carefully constructed for the cases studied, are limited in versatility and accuracy, while most of the existing first-principles band structures are computationally too demanding to study many realistic problems. The best—and a numerically tractable—first-principle theory is given by the local density approximation or LDA (Hohenberg and Kohn, 1964; Kohn and Sham, 1965; Baraff and Schluter, 1983).

One of the widely used methods to solve LDA equations is based on self-consistent pseudopotentials with plane-wave-basis expansions (Ihm and Cohen, 1980; Nielson and Martin, 1985; Ismail et al., 1991). The electronic and structural properties such as total energies, lattice constants, bulk moduli, and phonon frequencies are often in excellent agreement with experiments. Although this method has a built-in simplicity in constructing the Hamiltonian, the matrix size can be unmanageably large in studies, for example, of superlattices, defects, surfaces, and high-field transport that involve a large number of atoms. Other accurate methods that use linearized muffin-tin orbitals (Andersen, Jepsen, and Sob, 1987; van Schilfgaarde and Methfessel, 1990), and linearized, augmented plane waves (Zunger and Freeman, 1977; Mbaye, Ferreira, and Zunger, 1987) will also be time consuming, particularly in the studies listed above.

One way to overcome this hurdle is to use the linear combination of atomic orbitals (LCAO) method in a tight-binding (TB) form. This first-principles tight-binding (FPTB) Hamiltonian uses pseudo-atomic orbitals as the basis for a self-consistently determined Hamiltonian matrix. In a simple application, these orbitals are computed self-consistently from a free-atom calculation within LDA and used in the solid with a non-self-consistent Harris approximation (Harris, 1985). This method has been shown to be computationally fast and precise enough to be useful for a variety of problems (Jensen and Sankey, 1987, 1989; Sankey and Niklewski, 1989; Chelikowsky and Louie, 1984; Vanderbilt and Louie, 1984). A major advantage of the LCAO method over other methods is its availability in the familiar TB form which is perfectly suitable for all the tasks studied here. In particular, incorporation of this Hamiltonian in the Green's function approach whether to study surface-driven ordering, to obtain coherent potential approximation (CPA) limited alloy band structures, or to study the effect of impurities and defects on band structures, is straightforward.

2.1.2.1 LCAO Band Structures

In this section, we briefly describe the LCAO method. Details will be published elsewhere. By assuming only four orbitals (sp^3) per atom, we studied the band structures and structural properties of Group IV elements and their alloys, and of III-V compounds. Readily available pseudopotentials are used within LDA in the Kohn-Sham equation for individual atoms. A wave function obtained in this way at a site \mathbf{Q} is denoted as $|\alpha\mathbf{Q}j\rangle$, where α is s , p_x , p_y , or p_z , and j is anion or cation. In the solid, the wave functions, which are assumed to be a linear combination of these atomic wave functions, satisfy the Kohn-Sham (KS) eigenvalue equation

$$\left[\frac{p^2}{2m} + V_{\text{core}} + V_{\text{nonloc}} + V_{\text{coul}} + \mu_{\text{xc}} \right] \psi_i = \epsilon_i \psi_i \quad (1)$$

where i is the band index. The coulomb and exchange-correlation (xc) terms depend on the input electron density, $\rho_{\text{in}}(\mathbf{r})$ and the output charge density $\rho_{\text{out}} = 2\sum_i |\psi_i|^2$. The equation is solved for self-consistency (i.e., $\rho_{\text{in}} = \rho_{\text{out}}$); μ_{xc} is the total derivative of ρ time ϵ_{xc} with respect to ρ . Then the total energy is

$$E_{\text{tot}} = 2 \sum_i \epsilon_i + \frac{e^2}{2} \left[\sum_i \frac{Z_i Z_{i'}}{|\mathbf{r} - \mathbf{r}'|} - \int d\mathbf{r} \rho V_{\text{coul}} \right] + \int d\mathbf{r} \rho(\mathbf{r}) [\epsilon_{\text{xc}}(\rho) - \mu_{\text{xc}}(\rho)] \quad (2)$$

When the lattice periodicity exists, a Bloch state for a given crystal momentum, \mathbf{k} , is constructed as

$$|\alpha \mathbf{k} \mathbf{j}\rangle = \left[\frac{1}{N} \right]^{1/2} \sum_i e^{i\mathbf{k} \cdot (\mathbf{r}_i + \mathbf{r}_j)} |\alpha \mathbf{j}\rangle . \quad (3)$$

In addition, we expand the Bloch states in terms of plane waves, and we get

$$|\alpha \mathbf{k} \mathbf{j}\rangle = \sum_{\mathbf{G}} C_{\alpha}^j(\mathbf{k} + \mathbf{G}) |(\mathbf{k} + \mathbf{G})\rangle . \quad (4)$$

From Eqs. (1) and (2), we get the coefficients C_{α}^j in terms of the Fourier coefficients of the α orbital of the j atom. Once we obtain the Fourier transform of the kinetic energy and all potentials of the KS equation, then a \mathbf{k} -space 8×8 Hamiltonian is easily constructed.

As a test of accuracy of the results obtained with these LCAO Hamiltonian, we initially carried out the calculations with an assumption that the charge density in the solid is a superposition of atomic charge densities. This approximation, commonly known as the Harris approximation (Harris, 1985), removes the self-consistency loop and has been shown to yield reliable trends in bulk semiconductors. In this approximation, the Coulomb potential, V_{coul} , which depends on the total charge density, also reduces to the sum of contributions from individual atoms. However, the exchange correlation term is the most difficult one because of its complicated dependence on the charge density. A Fourier transform of the exchange correlation potential is obtained by direct space integration within a parallelepiped unit cell that contains two atoms. Ten special \mathbf{k} -points are used to calculate the E_{tot} as given in Eq. (2)

To study problems such as molecules, defects, and surfaces, it is necessary to have Hamiltonian matrix elements connecting any two orbitals in real space. For a periodic lattice, such a matrix can be obtained by simply integrating the above-calculated \mathbf{k} -space Hamiltonian (with an appropriate phase factor) over the Brillouin Zone (BZ). For a general case, we constructed the real-space Hamiltonian by expanding off-site functions in spherical harmonics defined with respect to a fixed origin. All integrals are then carried out in direct space. For the case with lattice periodicity, Hamiltonians obtained with these two methods have been tested and shown to agree with one another.

Band structure and total energy calculations have been carried out for Group IV elements, SiC, SiGe, and III-V compounds. The structural properties such as equilibrium bond length, d_0 ; cohesive energy per bond, E_b ; and bulk modulus, B_0 are obtained.

The results obtained for silicon are in excellent agreement with experimental results. The d_0 and B_0 were larger than the experimental values by 0.4% and 3%, respectively. The calculated cohesive energy per atom, -5.20 eV, agrees well with experiment (-4.64 eV) after subtracting the spin-polarization energy of -0.6 eV per atom. (The experimental ground state, as dictated by Hund's rule, has two p-electrons with the same spin, where the uncorrected calculations consider p-electrons with opposite spin).

The results obtained for other Group IV elements, such as germanium and carbon, are in reasonable agreement with experiments. For germanium, d_0 and B_0 differ from experiment by 0.4% and -11%, respectively, and E_b is 2.36 eV, which is larger than the experimental value of 1.94 eV. For carbon, d_0 and B_0 differ from experiments by -2.6% and 10%, respectively, and E_b is 3.99 eV, which is again larger than the experimental value of 1.68 eV.

Corresponding results for III-V compounds are considerably worse. In general, the d_0 is off by 2 to 8%, B_0 is off by 10 to 40%, and E_b is deeper by 0.2 to 0.9 eV. Because of the charge transfer from cation to anion, the wave functions in the solid are more localized than the corresponding atomic-wave functions. Linear combinations of atomic functions do not adequately describe the wave function in the solid. To improve matters, the basis functions are made more compact by multiplying with a function (van Schilfgaarde and Methfessel, 1990),

$$\frac{1}{(e^{\beta(r-r_0)} + 1)}$$

and renormalizing. The parameters β and r_0 are then varied to obtain a local maximum in the cohesive energy. As a general guiding rule, we start with twice the value of Pauling radius for r_0 . Values of 1 and 0.5 for β in carbon and germanium respectively, give excellent structural properties. Without further modifications, the total energy calculations are carried out for SiC and SiGe compounds. The calculated excess energy that favors ordering in SiC and segregation in SiGe alloys are in agreement (within a few millielectron volts) with experimental values.

Although such a (maximizing) procedure improved the accuracy, the disagreement was still very large for III-V compounds. We find that increasing r_0 and β reduces both the equilibrium bond length and the bulk modulus. However, when these parameters are varied slightly in the neighborhood of these 'equilibrium' values, excellent fits to experimental values are obtained for arsenide and phosphide compounds of germanium, aluminum, and indium.

2.1.2.2 Green's Function

In all the studies we carried out previously, we used Green's functions with an assumption that the basis set is orthonormal. However, the LCAO band structure described in the previous section has a basis set that is far from orthonormal. Hence, Green's function calculations have to be modified to the case of nonorthogonal basis orbitals. We define a Green's function G ,

$$(E + i\delta - H) G = 1 \quad (5)$$

When the basis set is orthogonal,

$$(E + i\delta) G_{\alpha\beta} - \sum_{\alpha'} H_{\alpha\alpha'} G_{\alpha'\beta} = \delta_{\alpha\beta} \quad (6)$$

then

$$G_{\alpha\beta} = \sum_i \frac{\langle \alpha | \psi_i \rangle \langle \psi_i | \beta \rangle}{E + i\delta - \epsilon_i} \quad (7)$$

where $H | \psi_i \rangle = \epsilon_i | \psi_i \rangle$. However, when the basis set is not orthonormal, Eq. (5) becomes

$$(E + i\delta) S_{\alpha\alpha'} G_{\alpha'\beta}^m - \sum_{\alpha'} H_{\alpha\alpha'} G_{\alpha'\beta}^m = S_{\alpha\beta} \quad (8)$$

It can be shown that all observables are related to G^m . For example, the density of states (DOS) is

$$\rho = -\frac{1}{\pi} \text{Im} \sum_{\alpha} G_{\alpha\alpha}^m \quad (9)$$

Note that $G_{\alpha\beta}^m$ is not a simple extension of Eq. (7). Now consider a Green's function G^k that satisfies

$$(E + i\delta) S_{\alpha\alpha'} G_{\alpha'\beta}^k - \sum_{\alpha'} H_{\alpha\alpha'} G_{\alpha'\beta}^k = \delta_{\alpha\beta} \quad (10)$$

Then

$$G^k = \frac{1}{ES - H + i\delta} \quad (11)$$

whose matrix elements are obtained as in Eq. (7). Now we can see from Eqs. (8) and (9),

$$G^m = G^k S \quad (12)$$

Hence the prescription to calculate the Green's function, when the basis set is nonorthogonal, is:

- (a) Obtain the Green's function in the usual way after replacing $E + i\delta$ by $ES + i\delta$. This is known as the contravariant Green's function, G^k .
- (b) Then multiply G^k by the overlap matrix S to obtain the mixed-mode Green's function, G^m . DOS is obtained from this Green's function as shown in Eq. (8).

After careful analysis, we are convinced that the difference equation method (see Section 2.3.1) we developed (Chen et al., 1989) can be modified to follow the above prescription. In addition to the H matrix, we have to calculate inter- and intralayer overlap matrices. Because the difference equation method is valid, calculation of surface and interface Green's functions are obtained without any difficulty. However, we find that the term values and interaction matrix elements of surface atoms are much different (as much as 2 eV in some cases) from their corresponding bulk values. In such cases, the surface DOS is obtained in two steps. First, by assuming the same term values on the surface, we use the difference equation method to obtain G^k . We then consider the difference in term values to obtain the final Green's function in a Dyson's expansion. Again, nonorthogonality requires some generalization and it has been included.

2.1.2.3 Generalized Perturbation Method

In the generalized perturbation method (GPM), the excess energy required to substitute CPA atoms by the constituent atoms is calculated from the Green's function obtained by Dyson's expansion (Turchi et al., 1987). Because of nonorthogonality, we must consider the change in the overlap matrix, in addition to the change in the potential. Once we obtain G^k with the Dyson's expansion, G^m is obtained by multiplying with the appropriate S matrix. The change in the DOS and energy are obtained from G^m in the usual way.

The existing LCAO Hamiltonian works well for Group IV elements and alloys (Krishnamurthy et al., 1992). However, the Hamiltonian for III-V compounds is better than the usual empirical TB Hamiltonian only in the sense that it has fewer (two) parameters and the spatial dependence of all matrix elements are known. The Harris approximation with only four

orbitals seems to break down for ionic compounds. For a truly first-principles LCAO Hamiltonian, we must remove the Harris approximation and obtain results self-consistently. Also, true atomic wave functions are more extended than required to describe the wave function in the solid. A systematic development of pseudo-atomic wave functions with d-orbitals is necessary for ionic compounds.

Our calculations of band structures suggests that interactions up to third neighbors should be included for reliable results. The existing Green's function programs consider only nearest bilayer plane interactions. For surfaces oriented in the (100) and (110) directions, this does not include all interactions up to third neighbors. We are generalizing the method by increasing the basic unit plane to include four monolayers. The ordering induced on these surfaces can then be studied.

Although the calculated valence bands agree very well with experiments, LDA does not produce the band gap nor some portion of the conduction bands accurately. The accurate evaluation of surface states may require accurate conduction bands. With a self-energy-like correction that contains the effect of exchange and correlation among electrons, known as GW approximation, a fully accurate band structure can be obtained (Hybertson and Louie, 1986, 1987, 1988). This self-energy term, which is nonlocal, energy dependent, and non-Hermitian, is evaluated from the dressed Green's function, G , and dynamical screening matrix.

2.1.3 Linear Muffin Tin Orbitals: Theory

In the past, all aspects of our Green's function and CPA work used empirical tight-binding methods. Although these methods are computationally easy to deal with, they suffer from a lack of general applicability and in many cases result in large quantitative uncertainties. However, much of our current and proposed work lies beyond the domain of simple tight-binding Hamiltonians; thus, our recent efforts have centered strongly around the use of an *ab initio* electronic structure method known as the linear muffin tin orbitals (LMTO) method. While we have recently obtained a number of results with this method (several are described below), we have until now mostly experimented with the various approximations within the method and their limitations.

Density-functional methods are quite powerful and remarkably accurate for a wide range of materials. Moreover, they are suitable for both electronic structure and mechanical properties. For example, native defect equilibrium has been studied in HgTe (Berding et al., 1992; Appendix B) which is an important feature in the performance of HgCdTe-based focal plane arrays. Their chief drawback has been that they are quite "heavy"; that is, they are difficult to implement and very slow to execute. The LMTO method is an efficient basis set; indeed, it is the most efficient of all electronic structure methods. Its primary drawback has lain in the spherical approximations made to the potential that render it unsuitable for calculating mechanical properties of structures of low symmetry. However, recent versions of the method remove that limitation.

We now possess three versions of the method.

- The first uses the classical atomic spheres potential approximation but is highly efficient; we have used it extensively in studies of, for example, Schottky barriers.

- The second uses a proper full-potential approximation, and thus solves the local-density functional nearly exactly. This method is still heavy and cumbersome, but is important to provide a reference for the correct local-density result.
- Finally, an efficient version of the second code implements the new "Harris-Foulkes" method (e.g., Foulkes and Haydock, 1989). The central idea here is to reevaluate the density-functional at the input density and make a sufficiently good guess for the density that it obviates the need for self-consistency. A superposition of spherical (e.g., free-atomic) densities seems to work remarkably well in almost all cases studied. Moreover, the resulting Hamiltonian looks quite similar to empirical tight-binding Hamiltonians. Should this method prove to be sufficiently accurate (as it seems to be), it should be an attractive alternative to the widely acclaimed Car-Parinello method (e.g., see Galli et al., 1989). The Harris functional, used in conjunction with the LMTO method (or some comparably efficient method) will be essential to the study of extended defects such as dislocations.

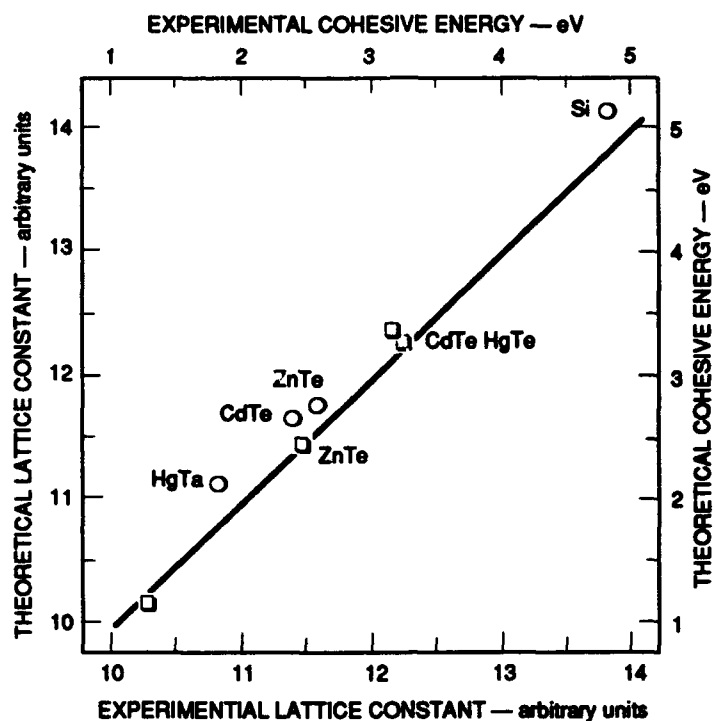
Figure 1 shows some results, in part to demonstrate the precision of the Harris method, but also with a view toward computing results of interest to us. Figure 1(a) compares the theoretical and experimental values of the lattice constant and cohesive energy; Figure 1(b) compares the theoretical and experimental values of the elastic constant, $c_{11} - c_{12}$, and the bulk modulus, B . Although the theoretical results calculated using the Harris method fare less well than do fully self-consistent calculations, the agreement with experiment is still quite good. Note that errors in the cohesive energy are due to problems with the local-density approximation, not due to Harris.

Figure 2 is closer to the practical applications our contract is aimed at. We seek a "supermodulus" effect in a thin superlattice of CdTe/ZnTe. In Figure 2, the superlattice consists of alternating monolayers, along the $\langle 110 \rangle$ direction, of CdTe and ZnTe. Interestingly, the bulk modulus is approximately linear (or sublinear) in the composition, while the shear modulus is supralinear. This shear modulus corresponds approximately to the measurements of Young's modulus by Farthing et al. (private communication).

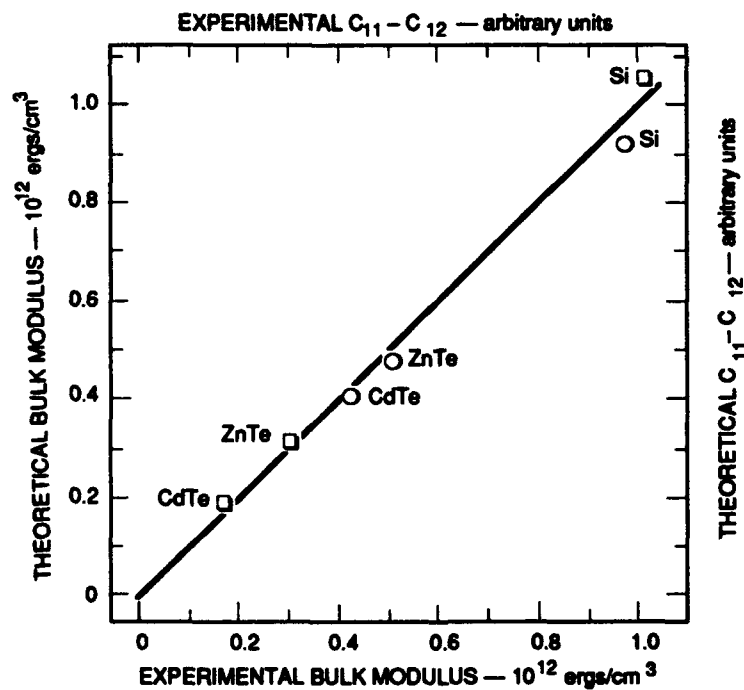
2.2 THEORY OF ELASTIC CONSTANTS

This work emphasizes three aspects: (1) analytical relations between elastic constants and microscopic atomic quantities, (2) quantitative calculation of elastic properties, and (3) extension to other mechanical properties. Most of our work concerns the first two aspects. Harrison's bond orbital model (BOM) provides a very good framework for the analytical work. This model treats the diamond and zinc blende semiconductors on equal footings. It does not need to treat the Madelung energies separately for polar systems, as required in the valence force field model. Although Harrison's universal parameters give only a qualitative prediction, the trends are correct. More important, these analytical expressions relate the three supposedly independent elastic moduli by the following formula:

$$9/C_{44} = 6/(C_{11} - C_{12}) + 4/B$$



(a) COMPARISON OF COHESIVE ENERGY (circles) AND LATTICE CONSTANT (squares) TO EXPERIMENTAL DATA



(b) COMPARISON OF $C_{11} - C_{12}$ (squares) AND BULK (circles) TO EXPERIMENT

32290-0

Figure 1. Cohesive energies and shear elastic coefficients: comparison between theory and experiment

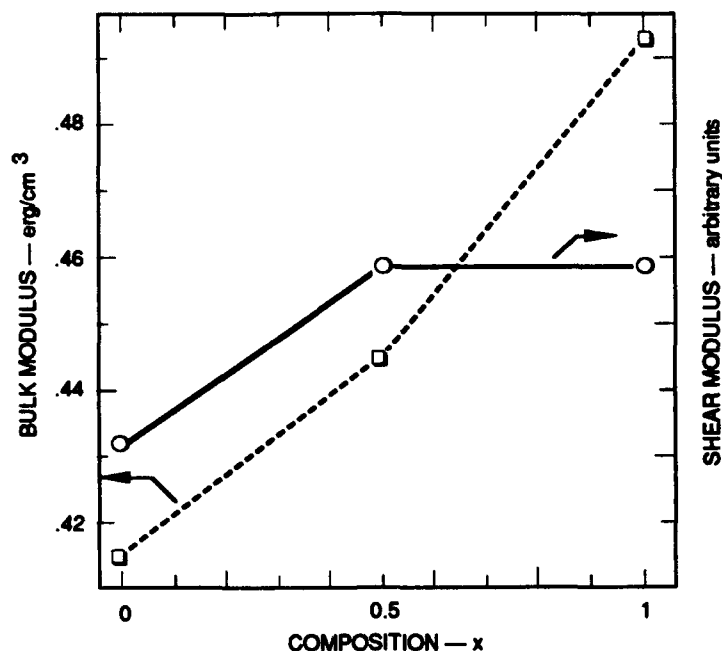


Figure 2 Calculation of shear modulus (circles) and bulk modulus (squares) for superlattices of CdZnTe along [110]. Shear compression is normal to [110] planes, keeping lateral area fixed.

where B is the bulk modulus. A test using the experimental values shows that this formula is satisfied to within 10% for all semiconductors. We have also modified the parameters to make the BOM a quantitative model.

For quantitative calculations, a tight-binding Hamiltonian similar to that used in BOM is adopted. The difference is that the Hamiltonian is diagonalized exactly in this approach whereas it is treated perturbatively and locally in BOM. There are four adjustable parameters: two (f and n) for the first-neighbor tight-binding Hamiltonian matrix elements, $V = f V_0 (d_0/d)^n$; and two (u_0 and m) for the repulsive pair energy, $u = u_0 (d_0/d)^m$. Here, V_0 is Harrison's universal matrix element, d is the distance between the nearest-neighbor atoms, and d_0 is its equilibrium value. These four parameters (f , n , u_0 , m) are adjusted so that the model produces the best experimental values for the bond length, cohesive energy, bulk modulus, and $C_{11} - C_{12}$. These parameters have been obtained for all common diamond and zinc blende semiconductors. The consistency of the model has also been confirmed by its ability to predict other elastic properties such as C_{44} and TO phonon frequencies (Yeh et al., 1991; Appendix C). The resulting Hamiltonian has also been used in other studies (Sher et al., Appendix D).

2.3 SURFACES, INTERFACES, AND SUPERLATTICES

2.3.1 Difference-Equation Approach

The difference-equation approach is a tight-binding (TB) based approach to layered structures without the need for using supercells. This method takes full advantage of the repeated units in these structures. The TB equations in the repeated regions are solved in terms of the

characteristic solutions to the difference equations which relate the expansion coefficients of the energy eigen wavefunction of a layer to those of the neighboring layers. The final problem is reduced to a small set of boundary equations. This approach also facilitates the calculation of Green's functions. Some analytical results for one dimension have been obtained and the theory has been generalized to three-dimensional systems. Several applications of this approach have also been made (Chen et al. 1989; Schick et al. 1989, Appendix E; Chen et al. 1990; Berding et al. 1990a).

2.3.2 Cleavage Energy

One application of the difference-equation approach is the calculation of the surface and cleavage energies in semiconductors. We applied this method to a representative group of semiconductors including silicon, GaAs, CdTe, and HgTe, using a second-neighbor tight-binding Hamiltonian. A Green's function method was used which exploits the symmetry of the surface problem and leads to a calculation of the surface energy directly, without requiring us to take the difference between two large numbers. We find $E\gamma(111) = 1360 \text{ ergs/cm}^2$ in GaAs and $E\gamma(110) = 1000, 180, \text{ and } 120 \text{ ergs/cm}^2$ in GaAs, CdTe, and HgTe, respectively. These results are in good agreement with the available experimental values. The results of this work have been published in the *Journal of Applied Physics* (Berding et al., 1990a).

The LCAO Green's function developed above for the surface was first used to calculate the cleavage energy for silicon whose cleavage plane is (111). The result was excellent bulk band structures of silicon using our FPTB Hamiltonian, which includes all interactions out to the third neighbor. Incidentally, when the planar unit in the difference equation approach is bilayered, all interactions up to third neighbor are included in the interplane Hamiltonian. In other words, previously written Green's function programs can be used with appropriate generalization for nonorthogonality and without further modification. The cleavage energy is calculated as described previously (Berding et al., 1990a). The calculated cleavage energy of 1260 ergs/cm^2 for silicon is in excellent agreement with experiment at $1140\text{--}1240 \text{ ergs/cm}^2$. Similarly, the calculated energy for GaAs agrees well previously reported value (Berding et al., 1990a). However, the (110) plane is the cleavage plane for GaAs, and such calculations have not yet been carried out for that plane. This is because not all third-neighbor interactions are included in the interbiplanar interaction for the (110) surface, and this requires some modification to existing Green's function programs. These modifications are under way.

2.3.3 Energetics of Atoms on Surfaces

A second-neighbor tight-binding model and the Green's function based on the difference equation approach in Section 2.3.1 were used to calculate the removal energies of constituent atoms from semiconductor surfaces. We found the energies from a fully covered surface to be substantially different from those from a nearly empty surface. These energies also vary from surface to surface and do not exhibit any simple relationship with the bulk cohesive energies. The energetics and a simple thermodynamic growth model are shown to explain anomalies in the MBE growth of HgCdTe. Similar approaches have also been used to calculate the cleavage

energies of semiconductors. The results compare well with available data. Several papers have been published on this subject (Krishnamurthy et al. 1990a, Krishnamurthy et al. 1990b, Krishnamurthy et al. 1990c).

We studied surface-induced ordering using the above-described procedure of surface and LCAO GPM Green's function. Although no ordering is expected on a (111)-oriented silicon surface, we carried out the calculations for that surface as a test of accuracy. The interatom interactions on the surface are strongly attractive so—in agreement with experiments—we predict growth with the formation of islands (i.e., two dimensional layer-by-layer growth). We carried out similar calculations on the (111)B surface of $\text{Ga}_{0.5}\text{Al}_{0.5}\text{As}$. The excess pair energy is zero within the accuracy of our calculations. However, the energy has to be less than zero (in our sign convention) to obtain ordering and indeed, experiments on the (111)B surface of this alloy do not find ordering. Ordering has been observed on (100) and (110) oriented surfaces of $\text{Ga}_{0.5}\text{Al}_{0.5}\text{As}$ alloys. Once our Green's function programs are extended to consider a four-layer unit plane, instead of the bilayer unit plane, we will be able to study phenomena related to other surfaces.

2.3.4 Surface Segregation in Alloys

Pseudo-binary semiconductor alloys almost always have a surface concentration that is different from the bulk in order to maintain a constant chemical potential for each layer in the alloy. We have calculated the degree of surface segregation in these alloys, with numerical examples presented for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ and $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$. The enthalpy responsible for segregation is the difference in the energies for moving an A or B atom from the bulk alloy to the surface. There are two major contributions to this energy process: (1) a bond-breaking mechanism, whereby the element with the lowest surface energy segregates to the surface; and (2) strain energy, where the dilute element in the compound is favored to segregate to the surface to alleviate the strain due to mismatch of the AC and BC bond lengths. We use the quasi-chemical approximation to calculate the free energy and hence the chemical potentials. Our results indicate that there is strong surface enrichment of mercury in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ while there is less surface segregation of mercury in $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ than the low x values appropriate for infrared applications. Mercury segregation to the surface will lower the bandgap or may turn the surface into a semimetal, thereby affecting the passivation of the surface. Several papers have been published on this subject (see, for example, Patrick et al. 1989, Berding et al. 1990c).

2.4 ORDER-DISORDER THEORY

2.4.1 Formation Energies and Phase Stability

The tight-binding model discussed above was used to calculate the formation energies, bond lengths, and bulk moduli of a number of III-V and II-VI semiconductor alloys. The input parameters are those of the constituent compounds. The model is then applied to alloys without further adjustment. We conclude that none of the ordered alloys found experimentally is in its stable bulk state at the growth temperatures (Yeh et al. 1991). Although the alloy excess energies can be negative, if the reference constituent compounds are constrained to match the substrate lattice used in the epitaxial growth, their magnitudes are not large enough to account for the observed ordering. We present a possible explanation of the observed states in terms of a

barrier that prevents the metastable ordered alloy from decomposing into separate phases. However, this explanation applies only to alloys with a large lattice mismatch between the constituent systems. The next step is to see if surface ordering is taking place and, if it is, what mechanism keeps the alloy from relaxing to the stable phases.

2.4.2 Critical Temperature of Order-Disorder Transitions

There is a large literature on order-disorder transition of alloys that ranges from semiempirical to highly theoretical models. The semiempirical models include the Bragg-Williams ideal solution models, Guggenheim's quasi-chemical approximation (QCA), the Ising model with empirical pair interactions, Stringfellow's δ -doping model, and Fedder and Muller's improvement of the δ -doping concept which removes some imprecision. All these methods miss portions of the problem: they either treat the excess energy of clusters in the pseudobinary alloy $A_{1-x}B_x$ or $A_{1-x}B_xC$ as being independent of concentration, or they make poor approximations to the combinatorial analysis of the number of permitted arrangements of atoms in the alloy. No theory has been published that relies on near-first-principles methods and proper combinatorics and that makes the relationship between all these methods and the true physics clear (and, more important, reaches proper conclusions). This material has been assembled into Chapter 4 of a book on semiconductor alloys being written by A.-B. Chen and A. Sher. A draft copy of Chapter 4 is appended to this report (Appendix F); we emphasize that the chapter is still undergoing revision.

2.4.3 Cluster Energies

We are recalculating the cluster energies that enter into the statistical mechanical calculation of correlations and phase diagrams. We are using the new parameterized tight-binding model discussed in Section 2.2.1. Previously, the 16-bond cluster energies were calculated from the sum of strain and chemical energies (Sher et al., 1987; Sher et al., 1991, Appendix G). Strain energies were calculated using Keating's force constant model, while chemical energies were calculated using Harrison's metallization terms within the extended bond-orbital model. In the present approach, we are using a cluster calculation similar to that used in computing the vacancy formation energies (Berding et al., 1990b), in which the 16-bond cluster Hamiltonian is directly diagonalized. We are using our empirical tight-binding Hamiltonian discussed in Section 2.2.1. In this cluster calculation, the positions of the twelve atoms at the edge of the cluster are held fixed while the positions of the five alloy atoms in the cluster are permitted to relax to the equilibrium position. Boundary conditions can be imposed by specifying the position of the atoms at the cluster edges. We are including the nonisotropic boundary conditions appropriate to the calculation clusters on a lattice-constraining substrate. Cluster energies will be used in the statistical mechanics to calculate correlations on epitaxial surfaces. Results of this work will be published upon completion.

2.5 DISLOCATIONS AND ELECTRO-OPTIC PROPERTIES

It is observed that dislocations in semiconductors, in particular GaAs and HgCdTe, have a significant impact on their electro-optic and transport properties. Both in GaAs at room temperature and $Hg_{0.8}Cd_{0.2}Te$ at 77 K, it is found that dislocations begin to modify the minority carrier lifetime, or equivalently the diffusion length, for dislocation densities in excess of

10^5 cm^{-2} , i.e., average separations of $\sim 30 \text{ }\mu\text{m}$. At higher densities, the lifetime decreases quadratically with the dislocation density for low carrier concentrations. This indicates that the dislocations are somehow combining to speed the recombination process. In GaAs, as the carrier concentration increases, starting from low values of $\sim 10^{15} \text{ cm}^{-3}$, the lifetime increases for a fixed dislocation density of $\sim 10^5 \text{ cm}^{-2}$; for high-enough concentrations, where Shockley Reed impurity recombination becomes dominant, lifetime decreases once more. Similar effects have been observed in HgCdTe.

In an attempt to understand the origin of these and other effects we have calculated the state densities in and near the bandgap of all four types of edge dislocations found in zinc blende structured semiconductors. The four types are those that terminate at cores on anion or cation lines either leaving one (the "shuffle" set) or three (the "glide" set) dangling bonds projecting into the core. Some of these dislocation types at 0 K exhibit partially filled states in the gap, and others have some filled conduction band states, or empty valance band states. Thus, those dislocations with partially filled gap states will be surrounded by depletion layers in both n- and p-type material, and the others will have either depletion or accumulation layers around them, depending on the carrier type of the host material. The behavior of these four different kinds of dislocations as they cross a p-n junction remains to be sorted out.

The dislocations with depletion layers serve as efficient recombination centers since their fields attract minority carriers into the core where recombination is fast. The effective capture cross section is high because, in the simplest picture its length scale is set by the depletion layer thickness. However, even this large length ($\sim 1 \text{ }\mu\text{m}$ in material with a low carrier concentration of $\sim 10^{15} \text{ cm}^{-3}$) is still small compared to the $\sim 30 \text{ }\mu\text{m}$ separation between dislocations when the dislocations begin to become effective lifetime killers.

The fields associated with a depletion layer, in this case around a line charge, do not fall abruptly at its edge. For a line charge, the field outside the depletion layer is asymptotically proportional to $\left(\frac{L_D}{r}\right)^{1/2} e^{-\frac{(r-R_D)}{L_D}}$ where r is the radial coordinate centered on the core in cylindrical coordinates, R_D is the dislocation radius, and L_D is the Debye length. (For a point charge screened by free carriers, the fields fall faster, as $\left(\frac{L_D}{r}\right)e^{-r/L_D}$.) If the distance a minority carrier drifts in these depletion layer fields in the minority carrier lifetime is comparable to the separation between dislocations, the dislocations become effective lifetime killers. We are still in the process of making accurate calculations, but our preliminary estimates indicate that at low carrier concentrations, where L_D is longest, the fields are sufficiently large in these high-mobility materials to become effective. Obviously they will be more effective in p-type material, where the minority species are the high-mobility electrons in the conduction band.

Given this mechanism, the reason the lifetime at first increases with carrier concentration becomes a natural consequence of the fact that L_D is inversely proportional to the square root of the carrier concentration.

We expect the nonlinear lifetime variation with dislocation density will also be a consequence of the long ranges of the fields, becoming most important when the average separation between dislocations approaches the order of the depletion layer radius. This study is continuing, and must be completed before our speculations can be placed on a firm-enough footing to warrant refereed publication.

3. FIRST-PRINCIPLES APPROACH TO THE PLASTIC PROPERTIES OF HIGH-TEMPERATURE ALLOYS

3.1 INTRODUCTION

With this end-of-year report, current funding for this program having ceased although we are at the entrance to potentially the most productive phase of this work, we also attempt some final conclusions. In retrospect, our work over the past two years has led us into some isolated fields such as the theoretical strength and the Jones theory of alloying. In parallel, however, we have been developing the methods and understanding to tackle our original long-term goals. Of Tasks 6 and 7 proposed in our extension of 16 March 1990, Task 7 is completed and Task 6 is as good as completed. We have also made major inroads into the program as it was envisioned developing into Years 3 and 4: we have made real progress toward our goal of alloy design from first-principles quantum mechanics. We look forward to continuing this work—hopefully under the aegis of AFOSR—in the future.

The topics in this section are of three categories.

- (i) The two items that complete Tasks 6 and 7.
- (ii) Our new work on the transition-metal trialuminides, which embodies a genuine attempt at alloy design and looks towards the development of our work in future.
- (iii) “Jones’s theory of the Hume-Rothery phases revisited” is the title of a new paper (a draft of which is included as Appendix H) written in collaboration with M. Methfessel and D.G. Pettifor. This work, although somewhat outside the purview of our original proposal, has a significant impact in theoretical alloy design.

We report on these topics in the following four sections.

3.2 THE NI-AL PHASE DIAGRAM

Task 7 was to compute the Ni-Al phase diagram using our total energy methods for calculating internal energy differences and a suitable approximate scheme for the entropy. This has been done in a collaboration between ourselves and a group at INPG, France. The work is to appear shortly in *Journal of Physics: Condensed Matter* and a copy is included as Appendix I.

3.3 PLANAR FAULT ENERGIES IN L_{12} ALLOYS

Calculations of planar fault energies (antiphase boundary or APB, complex stacking fault or CSF, and superlattice intrinsic stacking fault or SISF) are of central importance in theories of deformation in the high-temperature intermetallics, as was emphasized in our original proposal

and is still evident, even though new theories of anomalous yield have appeared since then (Dimiduk, 1991). Task 6 was to compute these energies in Ni₃Al with particular attention paid to the translation state of the defect. We say that this task is as good as completed because while we have the results now for Pt₃Al, those for Cu₃Au and Ni₃Al have not been tested for convergence with respect to supercell size. We *can* say, though, that in these two cases the errors are approximately 15% and that the defect energies are overestimated by this order of magnitude. So far we have comprehensive data only on (111) defects; the study will not be complete until the energies of (001) APBs are found, since it is the ratio of (001) to (111) APB energies that is of particular interest. We regret that we are not able to continue since, as we show below, the results are extremely encouraging. Because the work is incomplete and hence unpublished, we give a detailed account here.

Our approach closely follows that of Yamaguchi et al. (1981), who used symmetry arguments and classical pair potential simulations to deduce the shapes of energy vs. translation curves for planar defects in L1₂ intermetallics. They could not say how particular metals would behave but only describe generic features in the problem. We can now supply the missing link by doing explicit quantum mechanical calculations. We illustrate the procedure in Pt₃Al.

We first inquire how effective the Harris-Foulkes approximation is, as is our usual practice. We have calculated the shear constants C and C' and find good agreement with the results of Fu and Yoo using the full potential LAPW method (Fu and Yoo, 1989). We then compare the L1₂-D0₁₉ and L1₂-D0₂₂ energy differences, which we show in Table 1, both self-consistent and in the Harris-Foulkes approximation. For later reference, we show also our results for Cu₃Au.

Table 1
STRUCTURAL ENERGY DIFFERENCES IN PT₃AL AND CU₃AU
(mRy/atom)

	L1 ₂ -D0 ₁₉		L1 ₂ -D0 ₂₂	
	H-F	s-c	H-F	s-c
Pt ₃ Al	6.8	6.5	11.1	10.5
Cu ₃ Au	1.9	1.7		

Convinced that the Harris-Foulkes approximation is appropriate in Pt₃Al, we construct supercells containing APBs parallel to (111) planes and separated by 3, 6, and 9 layers of perfect crystal. We then calculate the APB energy in each cell to monitor the convergence with cell size. The results are shown in Table 2.

Even when the defects are separated by only three layers of crystal, the APB energy is well converged. We then use the three-layer supercells to calculate the energy of the SISF, which does not relax in plane (Yamaguchi et al., 1981) and the energy and translation state of the APB and CSF. All these defects are parallel to (111). Finally, we use a six-layer supercell to calculate the energy of the unrelaxed APB parallel to (001).

Our calculations in Ni₃Al and Cu₃Au are self-consistent using three-layer supercells. So far, we have not fully checked the convergence with respect to supercell size in Cu₃Au and

Table 2
APB ENERGY IN PT₃AL AS FUNCTION OF SUPERCELL SIZE

No. of Layers	γ_{APB} (J/m ²)
3	0.57
6	0.54
9	0.53

Ni₃Al. A six-layer self-consistent calculation of the unrelaxed APB in Ni₃Al gives an energy of 0.28 J/m² compared with 0.33 in the three-layer self-consistent calculation. Hence, we estimate that the defect energies are exaggerated by about 15% in Ni₃Al.

We now show the results that we have obtained so far. Each planar defect is characterized (Yamaguchi et al., 1981) by the fault plane (111) or (001), the fault vector \mathbf{p} , and the translation vector \mathbf{T} . Thus, the net displacement across the fault is $\mathbf{p} + \mathbf{T}$. Table 3 shows the characters of each fault studied; as shown, the translation state is characterized by the value of t . The CSF, when $t = -1$, reverts to a perfect crystal (Yamaguchi et al., 1981). As $t \rightarrow 1$, the APB becomes an SISF. In either case, $t = 0$ refers to the unrelaxed defect.

Table 3
CHARACTERS OF PLANAR FAULTS IN THE L₁₂ LATTICE

	Fault plane	\mathbf{p}	\mathbf{T}
SISF	(111)	$\frac{1}{3}[\bar{1}\bar{1}2]$	0
CSF	(111)	$\frac{1}{6}[\bar{1}1\bar{2}]$	$\frac{1}{6}[\bar{1}1\bar{2}]$
APB	(111)	$\frac{1}{2}[0\bar{1}1]$	$\frac{1}{6}[\bar{2}11]$
APB	(001)	$\frac{1}{2}[\bar{1}10]$	$\frac{1}{2}[001]$

In Table 4, we show planar fault energies, with relaxed values and translation states in those cases calculated so far. The only (001) APB energy so far found is for Pt₃Al, and this is an unrelaxed value.

In Figure 3, we show plots of fault energy against translation state, reminding the reader that the results for Ni₃Al and Cu₃Au are estimated to be exaggerated by something around 15%. More work is needed here using larger supercells.

We are deferring full discussion of these results until the study is completed. This is certainly the first time that such a detailed series of calculations has been attempted, and we have no doubt that the results, when completed, will represent a significant step forward in our understanding of the deformation of L₁₂ intermetallics. We advance some brief conclusions now, however, from the most obvious points to emerge from Table 4 and Figure 3.

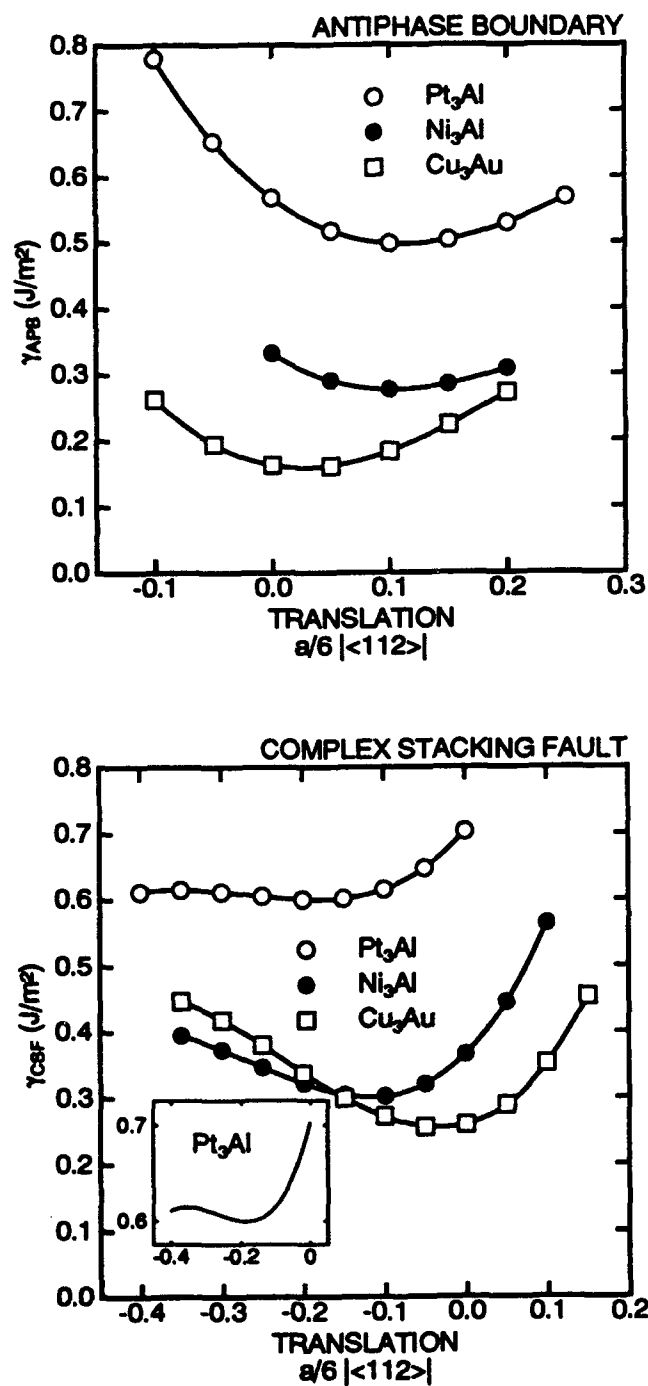


Figure 3 (111) planar fault energies in three L1₂ alloys as a function of in-plane translation. These local density calculations may be compared with similar curves obtained earlier using pair potentials (Yamaguchi et al., 1981). The near instability of the CSF in Pt₃Al can be seen in the inset which shows the curve on an expanded scale.

Table 4

PLANAR FAULT ENERGIES γ (J/m²) AND TRANSLATION STATES IN L1₂ INTERMETALLICS

	(111)					(001)	
	SISF	CSF		APB		APB	
	γ	γ	t	γ	t	γ	t
Pt ₃ Al	0.56	0.60	-0.181	0.50	0.112	0.54	—
Ni ₃ Al	0.05	0.30	-0.117	0.28	0.104		
Cu ₃ Au	0.12	0.26	-0.034	0.16	0.028		

We are deferring full discussion of these results until the study is completed. This is certainly the first time that such a detailed series of calculations has been attempted, and we have no doubt that the results, when completed, will represent a significant step forward in our understanding of the deformation of L1₂ intermetallics. We advance some brief conclusions now, however, from the most obvious points to emerge from Table 4 and Figure 3.

- In Pt₃Al, which is ordered to the melting point, the energies of the SISF and (001) APB, in which Al atoms are not forced into nearest neighbor positions, are comparable with the CSF and (111) APB in which there are Al-Al neighbors. This is surprising and contrary to the usual assumptions (Flinn, 1960). Results (indicated in Table 4) indicate that the SISF energy is close to zero in Ni₃Al, so that these two strongly ordering systems differ in this respect (it may be significant that Ni₃Al shows anomalous yielding while Pt₃Al does not).
- The CSF in Pt₃Al is practically unstable. γ_{CSF} has a minimum at $t = -0.181$ and a maximum at $t = -0.352$, the energy barrier being 0.015 J/m² corresponding to 0.24 kT per atom at the interface at room temperature. Thus, partial dislocations separated by CSF are not likely to be observed in Pt₃Al. In contrast, Ni₃Al and Cu₃Au, which both show anomalous yielding, have very stable CSFs.
- As expected from the weak ordering in Cu₃Au, the APB and CSF energies on (111) are lower than in Pt₃Al, and the translation states are smaller.
- As has been emphasized by Vitek (1990) but ignored in previous density functional calculations (Fu and Yoo, 1989), allowing the defects to relax can lead to significant changes in the calculated fault energy. In the extreme case of the CSF in Pt₃Al, relaxation causes more than 10% lowering. Now that we can predict the translation state, it becomes important to look for it experimentally in defect fringes and in variations in Burger's vector of dissociated dislocations.
- Finally, even allowing for overestimation of 15%, the calculated APB energy in Ni₃Al is significantly higher than any of the measured values, which are themselves rather scattered (Dimiduk, 1991). We need to investigate this discrepancy in future work. Maybe there is an error in the local density approximation, or maybe the theory and experiment can still be rationalized. We are not surprised, however, to

find disagreement between our results and a previous local density calculation (Fu and Yoo, 1989), since those authors used only very small supercells and calculated only two energies (γ_{SISF} and $\gamma_{\text{CSF}} + \gamma_{\text{APB}}$). As we see from our results, their claim that $\gamma_{\text{CSF}} = \gamma_{\text{APB}} + \gamma_{\text{SISF}}$ is greatly in error. Indeed, the equivalence is not even true in a pair potential description (Yamaguchi et al., 1981).

3.4 QUANTUM MECHANICAL ALLOY DESIGN AND THE TRANSITION-METAL TRIALUMINIDES

This work, undertaken in collaboration with D.G. Pettifor, will appear in the February 1992 issue of *Scripta Metallurgica* and a preprint is appended to this report (Appendix J). The paper indicates the direction in which our work is progressing toward the long-term goals outlined in our original proposal. A major aim is to use our first-principles theory to predict alloy compositions and stimulate experimental programs. Our reasons for choosing Al_3Ru as a likely candidate for a ductile transition-metal trialuminide are given in our paper. We have persuaded C.T. Liu of ORNL, Tennessee, to prepare some alloys and we look forward to the results of his investigations. So far, we understand, his results are promising. We have predicted the crystal structure of Al_3Ru to be D_{022} and not the reported D_{024} : in fact, as we point out, almost nothing is known of the equilibrium properties of Al_3Ru . We hope to work with the INPG group to predict the Al-Ru phase diagram, which is also known only tentatively.

3.5 THE JONES THEORY OF ALLOYING

Jones's early theory (1937) of the β -phase alloy of CuZn has had a turbulent history, and most textbooks are misleading, not to say wrong, on this subject. In collaboration with M. Methfessel and D.G. Pettifor, we have used modern electron theory and an advanced method of Brillouin zone integration to redo Jones's calculation using the rigid-band approximation with and without the inclusion of d electrons. From the band structure alone, we can now understand the origin of the face-centered cubic structure of the α -phase, and the stabilization of β - and ϵ -brass, as Cu is alloyed with Zn. It turns out that Jones's approach was valid but, although his conclusions were in perfect agreement with the facts, the agreement was entirely coincidental and due to the approximations he made. Without explicit inclusion of d electrons—which he, of course, neglected—one cannot predict all the phases of brass.

The writing-up of this work has taken some time, and led to a rather lengthy manuscript. A current draft is appended herewith.

REFERENCES

- Andersen, O.K., O. Jepsen, and M. Sob, 1987: *Electronic band structures and applications*, M. Yussouf, eds. (Springer Lecture Notes).
- Baraff, G.A., and M. Schluter, 1983: *Phys. Rev. B* **28**, 2296.
- Berding, M.A., M. van Schilfgaarde, and A. Sher, 1992: submitted to *J. Vac. Sci. Tech.*
- Berding, M.A., A. Sher, and A.-B. Chen, 1991: *Mat. Res. Soc. Symp. Proc.* **216**, 3.
- Berding, M.A., S. Krishnamurthy, A. Sher, and A.-B. Chen, 1990(a): *J. Appl. Phys.* **67**, 6175.
- Berding, M.A., A. Sher, and A.-B. Chen, 1990(b): *J. Appl. Phys.* **68**, 5064.
- Berding, M.A., A. Sher, A.-B. Chen, and R. Patrick, 1990(c): *Semicond. Sci. Technol.* **5**, S86.
- Chelikowsky, J.R., and S. Louie, 1984: *Phys. Rev. B* **29**, 2470.
- Chen, A.-B., Y.-M. Lai-Hsu, S. Krishnamurthy, M.A. Berding, and A. Sher, 1990: *Semicond. Sci. Technol.* **5**, S100.
- Chen, A.-B., Y.-M. Lai-Hsu, and W. Chen, 1989: *Phys. Rev. B* **39**, 923.
- Dimiduk, D.M., 1991: *J. Phys (Paris) III* **1**, 1025.
- Flinn, P.A., 1960: *Trans. TMS-AIME* **218**, 145.
- Fu, C.L. and M.H. Yoo, 1989: *Mat. Res. Soc. Symp. Proc.* **133**, 81.
- Galli, G., R.M. Martin, R. Car, and M. Parrinello, 1989: *Phys. Rev. Lett.* **62**, 555.
- Harris, J., 1985: *Phys. Rev.* **31**, 1770.
- Hohenberg, P.C., and W. Kohn, 1964: *Phys. Rev.* **136**, B864.
- Hybertson, M.S., and S.G. Louie, 1988: *Phys. Rev. B* **38**, 4033.
- Hybertson, M.S., and S.G. Louie, 1987: *Phys. Rev. Lett.* **58**, 1_51.
- Hybertson, M.S., and S.G. Louie, 1986: *Phys. Rev. B* **34**, 5390.
- Ihm, J., and M.L. Cohen, 1980: *Phys. Rev. B* **21**, 1527.
- Ismail, K., B.S. Meyerson, and P.J. Wang, 1991: *Appl. Phys. Lett.* **59**, 973.
- Jensen, R.W., and O.F. Sankey, 1989: *Phys. Rev. B* **39**, 3192.
- Jensen, R.W., and O.F. Sankey, 1987: *Phys. Rev. B* **36**, 6520.
- Kohn, W., and L. Sham, 1965: *Phys. Rev.* **140**, A1133.

- Krishnamurthy, S., A. Sher, and A.-B. Chen, 1992: in preparation
- Krishnamurthy, S., M.A. Berding, A. Sher, and A.-B. Chen, 1991: in *Computer Aided Innovation of New Materials*, M. Doyama, T. Suzuki, J. Kihara, and R. Áammamota, eds. (Elsevier Science Pub., North Holland), 681.
- Krishnamurthy, S., M.A. Berding, A. Sher, and A.-B. Chen, 1990a: *J. Appl. Phys.* **68**, 4020.
- Krishnamurthy, S., M.A. Berding, A. Sher, and A.-B. Chen, 1990b: *J. Cryst. Growth* **109**, 88.
- Krishnamurthy, S., M.A. Berding, A. Sher, and A.-B. Chen, 1990c: *Phys. Rev. Lett.* **64**, 2531.
- Krishnamurthy, S., A. Sher, and A.-B. Chen, 1989: *Appl. Phys. Lett.* **55**, 1002.
- Mbaye, A.A., L.G. Ferreira, and A. Zunger, 1987: *Phys. Rev. Lett.* **58**, 49.
- Nielsen, O.H., and R.M. Martin, 1985: *Phys. Rev. B* **32**, 3792.
- Patrick, R.S., A.-B. Chen, A. Sher, and M.A. Berding, 1989: *Phys. Rev. B*, 5980.
- Paxton, A.T., P. Gumbsch, and M. Methfessel, 1991: *Phil. Mag. Letts.* **63**, 267.
- Paxton, A.T., M. Methfessel, and H.M. Polatoglou, 1990: *Phys. Rev. B* **41**, 8127.
- Sankey, O.F., D.J. Niklewski, 1989: *Phys. Rev.* **40**, 3979.
- Schick, J.T., M. Bose, and A.-B. Chen, 1989: *Phys. Rev B* **40**, 7825.
- Sher, A., A.-B. Chen, and T. Yost, "Elastic Constants of Semiconductors and Alloys," to be published in *Semiconductors and Semimetals*.
- Sher, A., M.A. Berding, M. van Schilfgaarde, and A.-B. Chen, 1991: *Semicond. Sci. Technol.* **6**, C59.
- Sher, A., M.A. Berding, A.-B. Chen, and R. Patrick, 1989: *J. Cryst. Growth* **98**, 27.
- Sher, A., M. van Schilfgaarde, A.-B. Chen, and W. Chen, 1987: *Phys. Rev. B* **36**, 4279.
- Turchi, P., M. Sluiter, and D. de Fontaine, 1987: *Phys. Rev.* **36**, 3161.
- Vanderbilt, D., and S. Louie, 1984: *Phys. Rev. B* **30**, 6118.
- van Schilfgaarde, M., and M. Methfessel, 1990: *Bull. Am. Phys. Soc.* **35**, 667.
- Vitek, V., 1990: *Mat. Res. Soc. Symp. Proc.* **186**.
- Yamaguchi, M., V. Vitek, and D.P. Pope, 1981: *Phil. Mag. A* **43**, 1027.
- Yeh, C.-Y, A.-B. Chen, and A. Sher, 1991: *Phys. Rev. B* **43**, 9138.
- Yeh, C.-Y, A.-B. Chen, D.M. Nicholson, and W.H. Butler, 1990: *Phys. Rev. B* **42**, 10976.
- Zunger, A., and A.J. Freeman, 1977: *Phys. Rev. B* **15**, 5049.

Appendix A

**FULL-POTENTIAL KORRINGA-KOHN-ROSTOKER BAND THEORY APPLIED TO
THE MATHIEU POTENTIAL**

Full-potential Korringa-Kohn-Rostoker band theory applied to the Mathieu potential

Chin-Yu Yeh and A.-B. Chen

Physics Department, Auburn University, Auburn, Alabama 36849

D. M. Nicholson and W. H. Butler

Metals and Ceramics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

(Received 21 May 1990; revised manuscript received 8 August 1990)

The band theory of Korringa, Kohn, and Rostoker (KKR) based on the Green-function method is extended to space-filling potentials. A numerical test using the Mathieu potential shows good convergence for the bands up to 1.5 Ry with $l \leq 4$ included in the angular-momentum expansion for the wave functions. Our results strongly support the applicability of the full-potential KKR to bulk electronic-structure problems.

I. INTRODUCTION

The Korringa, Kohn, and Rostoker (KKR) band theory¹⁻² is an elegant theory for the one-electron energy bands in a closed-packed crystal for which the muffin-tin (MT) construction for the potential is a reasonable approximation. To expand the scope of application, considerable effort has been expended to extend the KKR theory to full crystal potentials.³⁻¹¹ One concern about such extension is related to the so-called near-field corrections (NFC) (Refs. 3-5) arising from the expansion of the KKR Green function beyond the muffin-tin region. Although there are proofs^{7,9,11} showing that NFC do not exist, questions have been raised about the applicability of the theory.¹² Since space-filling potentials are non-spherical and the Wigner-Seitz cell boundary is not smooth, we are further concerned about the speed of convergence in terms of angular-momentum (l) expansions. In this paper the integral equation approach of Kohn and Rostoker² (KR) is used to derive the full-potential KKR (FP-KKR) equation explicitly. One advantage of our derivation is that all the quantities involved are functions of r within a unit cell. Thus we can avoid the uncertainty in extending the wave function beyond the unit cell encountered in some other derivations.^{6,11} We have also tested the convergence by comparing the numerical results with the exact solution for the Mathieu potential¹³⁻¹⁵ in the simple-cubic crystal. Excellent results for the band structure in the energy range of interest are obtained with a maximum value of $l = 4$ included in this expansion.

The fact that the Mathieu potential is exactly soluble gives it an advantage for testing purposes over working with realistic potentials.^{4,9,17} Our test complements the empty-lattice potential¹⁸⁻²⁰ to provide a stringent test for the FP-KKR theory. The strong angular-momentum dependence in the Mathieu potential gives a good representation of the anisotropy that is present in the open structures pertaining to many semiconductors and insulators. The restriction of the KKR to closed-packed metals imposed by the muffin-tin approximation is lifted by

the full-potential method discussed here. The results obtained here should encourage the application of this theory to real crystals.

II. THE FULL-POTENTIAL KKR EQUATION

In this section we want to show that the Kohn-Rostoker integral equation can be simply extended to obtain the full potential KKR theory. The Schrödinger equation in the band calculation

$$[-\nabla^2 + V(r)]\psi_k(r) = E\psi_k(r) \quad (1)$$

for a full crystal potential $V(r)$ is equivalent to solving the following integral equation:²

$$\psi_k(E, r) = \int_{\tau} G_k(E; r, r') V(r') \psi_k(E, r') dr', \quad (2)$$

where the integration is over the Wigner-Seitz cell of volume τ , and k is a crystal wave vector. $G_k(E; r, r')$ in Eq. (3) is the KKR free-electron Green's function²

$$G_k(E; r, r') = -\frac{1}{\tau} \sum_n \frac{\exp[i(\mathbf{K}_n + \mathbf{k}) \cdot (\mathbf{r} - \mathbf{r}')] }{(\mathbf{K}_n + \mathbf{k})^2 - E}, \quad (3)$$

where \mathbf{K}_n are the reciprocal-lattice vectors. Alternatively $G_k(E; r, r')$ can be expressed as²

$$G_k(E; r, r') = -\frac{1}{4\pi} \sum_s \frac{\exp(i\kappa|\mathbf{r} - \mathbf{r}' - \mathbf{R}_s|)}{|\mathbf{r} - \mathbf{r}' - \mathbf{R}_s|} \exp(i\mathbf{k} \cdot \mathbf{R}_s), \quad (4)$$

where $\kappa = \sqrt{E}$ for $E > 0$ and $\kappa = i\sqrt{|E|}$ for $E < 0$, and \mathbf{R}_s are the lattice translation vectors. To derive the FP-KKR equation, we first observe that Eq. (3) can be cast into a surface integral,

$$\int_{S_r} [G_k(E; r, r') \nabla' \psi_k(E, r') - \psi_k(E, r') \nabla' G_k(E; r, r')] \cdot d\mathbf{S}' = 0, \quad (5)$$

where S_r is the surface of the Wigner-Seitz cell.

Since the $|\mathbf{r}'|$ in the surface integral exceeds the muffin-tin radius r_m , we need to consider the expansion of

the Green function beyond the original range of Kohn and Rostoker. Several authors have already considered this point. For simplicity, we shall only consider the case with one atom per unit cell. Since this expansion is a cen-

tral point of controversy, we rederive the results explicitly in the Appendix for the range of r and r' needed here. We show that the expansion

$$G_L(E; r, r') = \sum_L \left[\sum_{L'} [i^{(l-l')} B_{LL'}(\mathbf{k}, E) J_L(\kappa r) J_L'(\kappa r') + \kappa \delta_{LL'} J_L(\kappa r) N_L(\kappa r')] \right] \quad (6)$$

is valid as long as both r and r' are inside τ and satisfy the following condition:

$$|r| < |r'| < |R_s| \text{ for all } |R_s| \neq 0. \quad (7)$$

In Eq. (6) the notations $J_L(\kappa r) = j_l(\kappa r) Y_L(r)$ and $N_L(\kappa r) = n_l(\kappa r) Y_L(r)$ are used, where j_l and n_l are, respectively, the spherical Bessel and Neumann functions, Y_L is a real spherical harmonics, and L represents the double indices (l, m) . $B_{LL'}(\mathbf{k}, E)$ is the usual KKR structure constant.^{2,13} We note that for any $|r|$ smaller than r_m , the condition in Eq. (7) is satisfied for all r' contributing to the surface integration in Eq. (5). The condition $|r'| < |R_s|$ in Eq. (7) holds for most lattices; exceptions are those, for example, with long narrow cells. For such cases, this condition can be satisfied by breaking the unit cell into smaller cells including so-called "empty cells" which do not contain an atomic nucleus.

The wave function inside the cell τ can be expanded in a basis set $\{\Phi_L(E, r)\}$ as

$$\psi_L(E, r) = \sum_L a_L(\mathbf{k}, E) \Phi_L(E, r). \quad (8)$$

The basis function $\Phi_L(E, r)$ is a regular solution to the Schrödinger equation inside τ ,

$$[-\nabla^2 + V(r)] \Phi_L(E, r) = E \Phi_L(E, r), \quad (9)$$

and behaves like $J_L(\kappa r)$ at the origin $r = r_0 \rightarrow 0$, which is typically the location of the atomic nucleus. This basis set can be calculated using the following integral equation:⁴

$$\Phi_L(E, r) = J_L(\kappa r) + \sum_{L'} \int_{r_0}^r g_{LL'}(E; r, r') V(r') \Phi_{L'}(E, r') d^3 r', \quad (10)$$

where $g_L(E; r, r')$ is a free-particle Green's function and is defined as

$$g_L(E; r, r') = \kappa [J_L(\kappa r) N_L(\kappa r') - N_L(\kappa r) J_L(\kappa r')]. \quad (11)$$

We note that the basis function $\Phi_L(E, r)$ is coupled to other angular-momentum channels for $r > r_0$, because the crystal potential $V(r)$ is not spherical.

The expansions of G in Eq. (6) and of ψ in Eq. (8) can be substituted in Eq. (5) to obtain

$$\sum_L J_L(\kappa r) \sum_{L'} \left[\left[\sum_{L''} [i^{(l-l'')} B_{LL''}(\mathbf{k}, E) S_{L'L''}(E)] + \kappa C_{LL''}(E) \right] a_{L''}(\mathbf{k}, E) \right] = 0, \quad r < r_m, \quad (12)$$

where

$$S_{L'L''}(E) = \kappa \int_{S_\tau} [J_L(\kappa r'), \Phi_{L''}(E, r')] \cdot d\mathbf{S}', \quad (13)$$

and

$$C_{LL''}(E) = \kappa \int_{S_\tau} [N_L(\kappa r'), \Phi_{L''}(E, r')] \cdot d\mathbf{S}'. \quad (14)$$

In the above, the notation $[F_1, F_2] \equiv F_1 \nabla' F_2 - F_2 \nabla' F_1$ has been used. The surface integrals in Eqs. (13) and (14) are over the boundaries of τ as indicated by S_τ . Since $J_L(\kappa r)$ in Eq. (12) are linearly independent functions, the following set of homogeneous equations holds:

$$\sum_{L''} \left[\left[\sum_{L'} [i^{(l-l')} B_{LL'}(\mathbf{k}, E) S_{L'L''}(E)] + \kappa C_{LL''}(E) \right] a_{L''}(\mathbf{k}, E) \right] = 0. \quad (15)$$

This is the FP-KKR equation that we are after.

We note that our derivation is similar to Nesbet's derivation.¹¹ We hope, however, that the above explicit derivation may be more accessible to some readers. It is also useful for establishing the notation necessary for the description of the application of FP-KKR theory to the solution of the Mathieu potential which constitutes the main result of this paper.

III. CALCULATION OF S AND C MATRICES

The surface integrals for the S and C matrices in Eqs. (13) and (14) can be very time consuming. It is desirable to seek simplification of these calculations in a real application. One plausible approximation which is consistent with the KKR spirit is to expand every quantity involved in angular-momentum components. Equations (13) and (14) are equivalent to the volume integrations

$$S_{L'L''}(E) = \kappa \int_\tau J_L(\kappa r) V(r) \Phi_{L''}(E, r) d\mathbf{r} \quad (16)$$

and

$$C_{L'L}(E) = -\delta_{L'L} + \kappa \int_{\tau} N_{L'}(\kappa r) V(r) \Phi_L(E, r) dr. \quad (17)$$

One can free the limits of these integrals by replacing the crystal potential V by the truncated potential V^T ,

$$V^T(r) = V(r)\sigma(r), \quad (18)$$

where $\sigma(r)$ is a step function and is defined as

$$\sigma(r) = \begin{cases} 1, & \text{for } r \text{ within } \tau \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

The angular-momentum expansion for the basis function is assumed to be

$$\Phi_L(E, r) = \sum_{L'} \phi_{L'L}(E, r) Y_{L'}(r), \quad (20)$$

and the truncated potential V^T is expanded as

$$V^T(r) = \sum_L V_L^T(r) Y_L(r). \quad (21)$$

The integrations in Eqs. (16) and (17) can be reduced, respectively, to the simple radial integrations

$$S_{L'L}(E) = \kappa \sum_{L''} \int_{r_0}^{\tau} j_{L'}(\kappa r) V_{L'L''}(r) \phi_{L''L}(E, r) r^2 dr \quad (22)$$

and

$$C_{L'L}(E) = -\delta_{L'L} + \kappa \sum_{L''} \int_{r_0}^{\tau} n_{L'}(\kappa r) V_{L'L''}(r) \times \phi_{L''L}(E, r) r^2 dr, \quad (23)$$

where r_s is the radius of the circumscribing sphere of the Wigner-Seitz cell. The $V_{L'L''}(r)$ is given by

$$\begin{aligned} V_{L'L''}(r) &= \int Y_{L'}(r) V^T(r) Y_{L''}(r) d\Omega \\ &= \sum_L C_{L'L''}^L V_L^T(r), \end{aligned} \quad (24)$$

where

$$C_{L'L''}^L \equiv \int Y_{L'}(r) Y_L(r) Y_{L''}(r) d\Omega \quad (25)$$

is a Gaunt coefficient.

Note that in the above the basis function $\Phi_L(E, r)$ is assumed to be calculated from Eq. (10), where $V(r)$ is the full crystal potential. This is the same procedure used by Brown and Ciftan (BC).⁶ The original Williams-Morgan³ (WM) approach, however, used the truncated potential V^T for the calculation of the basis function in Eq. (10). If expansions of the potential and Φ_L in Eq. (10) include all the angular-momentum components, both approaches probably will give the same results for the band structure, provided both converge.^{11,21} In practice, the expansion is limited to a certain l_{\max} ; therefore, these two approaches yield different results.

In the actual calculation of the basis functions using either V or V^T , we first write the potential as the sum of V_0 and ΔV , where V_0 is the spherical part of the potential and ΔV is the rest. We then solve for the radial wave function f_l corresponding to V_0 . Similarly, the basis is written as $\Phi_L = F_L + \Delta\Phi_L$, where $F_L = f_L Y_L$, and $\Delta\Phi_L$ is solved from the integral equation

$$\begin{aligned} \Delta\Phi_L(E; r) &= \sum_{L'} \int_{r_0}^{\tau} g_{L'L}(E; r, r') \Delta V(r') F_L(E; r') dr' \\ &+ \sum_{L'} \int_{r_0}^{\tau} g_{L'L}(E; r, r') V(r') \Delta\Phi_L(E; r') dr', \end{aligned} \quad (26)$$

iteratively using angular-momentum expansions for all quantities involved.

IV. APPLICATION TO MATHIEU POTENTIAL

To test the accuracy of the FP-KKR equation and the convergence in angular expansion described above, we applied the theory to the Mathieu potential¹³⁻¹⁶ of the form

$$V(r) = -U_0 \left[\cos \frac{2\pi x}{\alpha} + \cos \frac{2\pi y}{\alpha} + \cos \frac{2\pi z}{\alpha} \right], \quad (27)$$

where we took the lattice constant α to be 2π times the Bohr radius and the potential parameter U_0 to be 0.5 Ry. Because the potential is separable, the eigenvalue problem reduces to three one-dimensional problems. The band structures and corresponding wave functions can be computed to the precision of the computer and can be regarded as "exact" in the numerical comparison.

The Mathieu potential is poorly represented by the MT approximation, because the simple cubic structure is rather open and the potential has a large variation in the

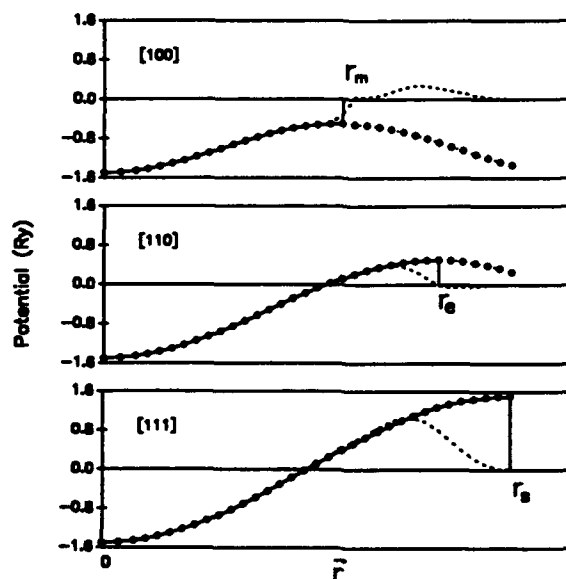


FIG. 1. Angular-momentum expansion of the Mathieu potential along [100], [110], and [111]. The solid circles represent the continuous crystal potential $V(r)$, and the solid lines are the truncated potential $V^T(r)$. The dotted and the dashed lines are the sums of the angular-momentum components up to $l=8$ for $V(r)$ and $V^T(r)$, respectively. r_m , r_e , and r_s are the distances between the origin and the face, edge, and corner of the cube, respectively. Notice that the dotted line and the solid circles are not distinguishable in the figure.

interstitial region. For example, with U_0 set to be 0.5 Ry, the MT constant potential is $V_c = U_0[9/\pi(6-\pi)] \approx 0.501116$ Ry, while the actual value of the potential varies from -0.5 Ry at $(\frac{1}{2}, 0, 0)\alpha$ to 1.5 Ry at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})\alpha$.

When the full potential in Eq. (27) is expanded in cubic harmonics $V(r) = \sum_L V_L(r)K_L(r)$, $V_L(r)$ is proportional to $-U_0 j_l(2\pi r/\alpha)$, and the series converges very fast. With an $l_{\max} = 8$, one can achieve a converged $V(r)$, as shown in Fig. 1. However, in the expansion for the truncated potential, $V^T(r) = V(r)\sigma(r) = \sum_L V_L^T(r)K_L(r)$, the components

$$V_L^T(r) = \int K_L(r)V(r)\sigma(r)dr \quad (28)$$

have to be carried out numerically with great care. Be-

cause of sharp edges and corners in $V^T(r)$, the angular-momentum expansion is only slowly converging. This is evident in Fig. 1, which shows sizable errors made in all three directions [100], [110], and [111] in the expansion of V^T up to $l_{\max} = 8$.

We have carried out the FP-KKR calculation using the wave-function expansion in Eqs. (10) and (24) up to $l_{\max} = 4$. The basis sets are calculated using both the BC and WM approaches with the potentials expanded up to $l_{\max} = 8$. Results from the MT-KKR approximation are also obtained for comparison.

In Fig. 2(a), the solid lines represent the "exact" band structures for the Mathieu potential. The dots are the MT-KKR results. Despite the crude approximation in the MT potential, the lowest band is still reasonable. The MT approximation becomes worse at the higher energies,

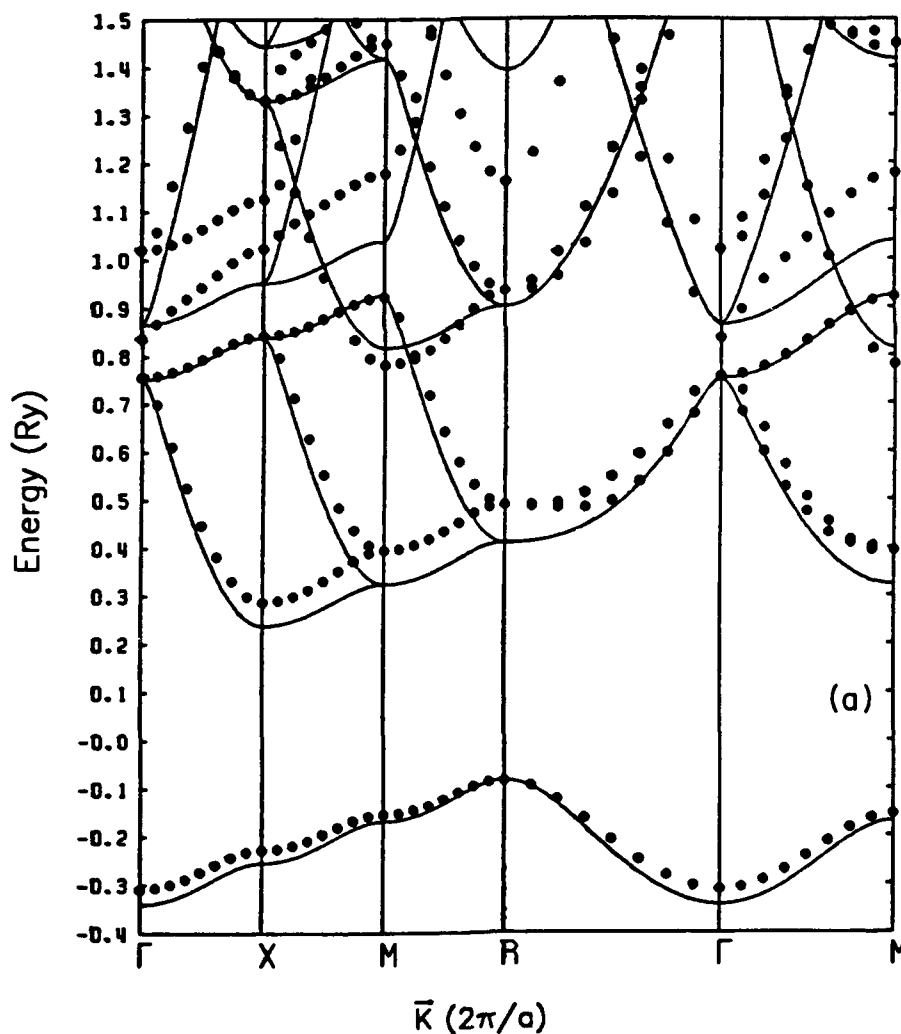


FIG. 2. Comparison of (a) the muffin-tin KKR and (b) the FP-KKR band structures (the dots) with the exact results (the solid lines) for the Mathieu potential. The symmetry points Γ , X , M , and R correspond to the wave vector at $(0,0,0)$, $(\frac{1}{2}, 0, 0)$, $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, respectively, in units of $2\pi/a$.

as indicated by the large energy deviations and splittings of the levels. For example, the "exact" bands from Γ to X from Γ to M around 0.9 Ry are degenerate due to separability of the Mathieu potential, while the MT approximation lifts this "accidental" degeneracy.

The full-potential KKR results are compared with the "exact" band structure in Fig. 2(b). The dots are now the FP-KKR results and are calculated based on the BC approach. The agreement is excellent and rather uniform up to 1.3 Ry. The calculation even preserves the accidental degeneracy at Γ at energy 0.88 Ry. The lowest band has a detectable deviation of 0.016 Ry at R , but has very small root-mean-square (rms) deviation. The deviations at R and some other energy states are probably due to the truncation in the angular-momentum expansions. The FP-KKR bands based on the WM approach are not no-

ticeably different from those based on the BC approach plotted in Fig. 2(b). However, there are slight differences between the results of the two approaches. For reference, we list the deviations of both the BC and WM approaches and the "exact" energies in Table I. While the WM approach gives a larger deviation in the lowest-energy band around R , the overall rms deviations of these two approaches are similarly small. These results imply some freedom in the choice of basis functions. Provided that reasonable approximations are made in the representation of the cell potential V^T and in the calculation of S_{LL} and C_{LL} from Eqs. (21) and (22), it appears that the FP-KKR equation will give reasonable bands independent of the exact algorithm for obtaining the Φ_L , e.g., from $V(r)$, $V^T(r)$, or other smooth potentials augmented to V^T .

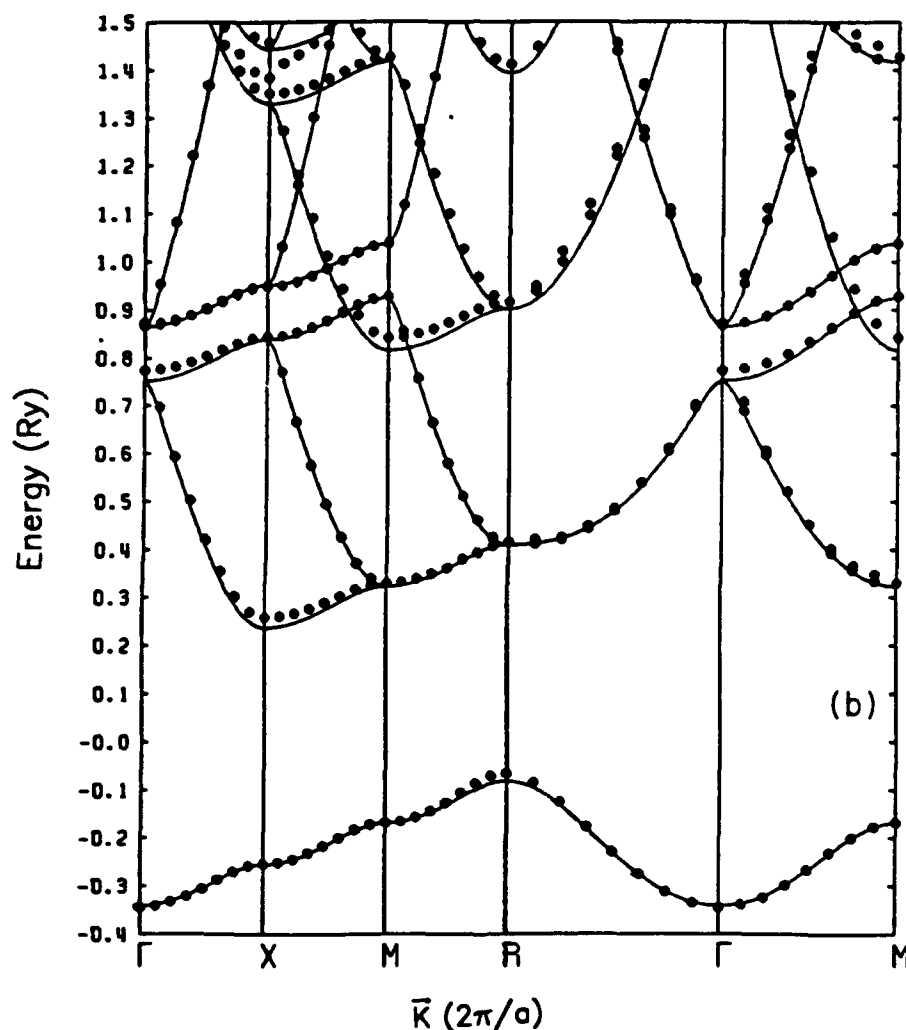


FIG. 2. (Continued).

TABLE I. Deviations of the FP-KKR band energies ΔE from the exact values E_{exact} for the Mathieu potential at several symmetry points. The subscripts WM and BC stand, respectively, for the Williams-Morgan and Brown-Ciftan approaches described in the text. All energies are in Ry.

Symmetry states	ΔE_{WM}	ΔE_{BC}	E_{exact}
Γ_1	-0.0053	-0.0034	-0.3414
Γ_{15}	0.0243	0.0223	0.7517
Γ_1	-0.0015	0.0024	0.8653
Γ_{12}	0.0034	0.0069	0.8653
R_1	0.0274	0.0160	-0.0827
R_{15}	0.0089	0.0064	0.4097
R_{25}	0.0156	0.0145	0.9020
R_2	0.0184	0.0171	1.3943
X_1	0.0024	0.0005	-0.2551
X_4	0.0214	0.0216	0.2372
X_5	0.0060	0.0056	0.8379
X_1	-0.0010	-0.0021	0.9515
X_2	-0.0047	-0.0036	0.9515
M_1	0.0016	0.0002	-0.1689
M_5	0.0072	0.0067	0.3234
M_3	0.0270	0.0259	0.8158
M_4	0.0033	0.0037	0.9241
M_1	0.0058	-0.0005	1.0377

V. SUMMARY

The main purpose of this paper is to test the accuracy of the FP-KKR theory in band-structure calculations. To help eliminate doubts about this theory, we have derived the FP-KKR equation explicitly from the Kohn-Rostoker integral equation.² This FP-KKR theory still preserves the clear separation between the structural and potential information possessed in the MT-KKR equation. The potential information is contained in the S and C matrices, which can be easily calculated if the basis functions and potentials are expressed in angular-momentum expansions. Such expansions are desirable in a realistic calculation. The whole procedure has been tested against the exactly soluble Mathieu potential in the simple-cubic structure. Because of the openness of the structure and the high anisotropy of the potential, this potential provides a challenging model to test against any band-structure theory. Our results show that with wave

functions expanded up to $l_{\text{max}}=4$ and the potential up to $l_{\text{max}}=8$, the FP-KKR theory as described above gives excellent results for the bands in the energy range needed for solid-state applications. With this method, one should be able to deal with solids having open structures, such as semiconductors, for which MT-KKR is not suitable.

ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with Dr. B. Segall, Dr. J. S. Faulkner, and Dr. A. Gonis. The research at Auburn University was supported in part by Office of Naval Research (ONR) Contract No. N00014-88-C0096 and Air Force Office of Scientific Research (AFOSR) Contract No. F49620-88-K-0009. The calculation was done by using the Alabama Supercomputer Network. The work at Oak Ridge National Laboratory was sponsored by the U.S. Department of Energy Division of Materials Sciences, Office of Basic Energy Sciences through Contract No. DE-AC05-8400R-21400 with Martin Marietta Energy Systems, Inc.

APPENDIX: GREEN-FUNCTION EXPANSION

Here we want to show that Eq. (6) is valid when Eq. (7) is satisfied. Following Kohn and Rostoker,² we separate G_k of Eq. (5) into two parts,

$$G_k(E; \mathbf{r}, \mathbf{r}') = g_0(\mathbf{k}, E; \mathbf{r}, \mathbf{r}') + g_1(\mathbf{k}, E; \mathbf{r}, \mathbf{r}'), \quad (\text{A1})$$

where g_0 is the singular part,

$$g_0(\mathbf{k}, E; \mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi} \frac{\exp(i\kappa|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}, \quad (\text{A2})$$

and

$$g_1(\mathbf{k}, E; \mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi} \sum_{s \neq 0} \frac{\exp(i\kappa|\mathbf{r}-\mathbf{r}'-\mathbf{R}_s|)}{|\mathbf{r}-\mathbf{r}'-\mathbf{R}_s|} \exp(i\mathbf{k} \cdot \mathbf{R}_s). \quad (\text{A3})$$

For $r < r' < R_s$ and for \mathbf{r} and \mathbf{r}' inside τ , the first part has the expansion $g_0 = -i\kappa \sum_L J_L(\kappa\tau) H_L^\dagger(\kappa\tau')$, where $H_L^\dagger(\kappa\tau) = J_L(\kappa\tau) + iN_L(\kappa\tau)$. Under the same condition for \mathbf{r} and \mathbf{r}' , $|\mathbf{r}| < |\mathbf{r}'+\mathbf{R}_s|$ also holds for a Wigner-Seitz cell, so that the following expansion is valid:^{6,8}

$$\begin{aligned} \frac{\exp(i\kappa|\mathbf{r}-\mathbf{r}'-\mathbf{R}_s|)}{|\mathbf{r}-\mathbf{r}'-\mathbf{R}_s|} &= -i\kappa \sum_L J_L(\kappa\tau) H_L^\dagger(\kappa(\tau'+R_s)) \\ &= -i\kappa \sum_L \left[\sum_{L'} J_{L'}(\kappa\tau) \mathcal{H}_{LL'}^\dagger(\kappa\mathbf{R}_s) J_{L'}(\kappa\tau') \right], \end{aligned} \quad (\text{A4})$$

where

$$\mathcal{H}_{LL'}^\dagger(\kappa\mathbf{R}_s) = 4\pi \sum_{L_2} i^{(l'+l_2-1)} C_{LL_2}^{L_2} H_{L_2}^\dagger(\kappa\mathbf{R}_s), \quad (\text{A5})$$

where $C_{LL_2}^{L_2}$ is given in Eq. (25). Therefore the Green function has the expansion in Eq. (6) with the structure constant given by

$$B_{LL}(\mathbf{k}, E) = -i\kappa \left[\delta_{LL} + i^{(l'-1)} \sum_{s \neq 0} \mathcal{H}_{LL}^\dagger(\kappa\mathbf{R}_s) \exp(i\mathbf{k} \cdot \mathbf{R}_s) \right]. \quad (\text{A6})$$

- ¹J. Korrington, *Physica* **13**, 392 (1947).
²W. Kohn and N. Rostoker, *Phys. Rev.* **94**, 1111 (1954).
³A. R. Willaims and J. van W. Morgan, *J. Phys. C* **7**, 37 (1974).
⁴J. S. Faulkner, *Phys. Rev. B* **19**, 6186 (1979).
⁵J. S. Faulkner and T. P. Beaulac, *Phys. Rev. B* **26**, 1597 (1982).
⁶R. G. Brown and M. Ciftan, *Phys. Rev. B* **27**, 4564 (1983).
⁷A. Gonis, *Phys. Rev. B* **33**, 5914 (1986).
⁸R. Zeller, *J. Phys. C* **20**, 2347 (1987).
⁹A. Gonis, X.-G. Zhang, and D. M. Nicholson, *Phys. Rev. B* **38**, 3564 (1988); **40**, 947 (1989).
¹⁰X.-G. Zhang and A. Gonis, *Phys. Rev. B* **39**, 10373 (1989).
¹¹R. K. Nesbet, *Phys. Rev. B* **41**, 4948 (1990).
¹²E. Badraxe and A. J. Freeman, *Phys. Rev. B* **36**, 1378 (1988).
¹³F. S. Ham and B. Segall, *Phys. Rev.* **124**, 1786 (1961).
¹⁴B. Segall, *J. Phys. Chem. Solids* **8**, 371 (1959).
¹⁵P. M. Morse, *Phys. Rev.* **35**, 1310 (1930); *Tables Relating to Mathieu Functions* (Columbia University, New York, 1951).
¹⁶J. C. Slater, *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York, 1965), Vol. 2, pp. 158-166.
¹⁷D. M. Nicholson and J. S. Faulkner, *Phys. Rev. B* **39**, 8187 (1989).
¹⁸R. Zeller, *Phys. Rev. B* **38**, 5993 (1988).
¹⁹J. S. Faulkner, *Phys. Rev. B* **32**, 1339 (1985).
²⁰R. G. Brown and M. Ciftan, *Phys. Rev. B* **39**, 3543 (1989).
²¹Reference 11 showed that the two procedures led to the same basis functions Φ_L .

Appendix B
DEFECT EQUILIBRIUM IN HgTe

Defect equilibrium in HgTe

M.A. Berding, M. van Schilfgaarde, and A. Sher
SRI International
333 Ravenswood Ave., Menlo Park, California 94025

ABSTRACT

Defects are a well known source of problems in low x $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ and $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$. To gain insight into the relative importance of various native point defects in these materials we have calculated the formation energy of native point defects in HgTe using a full potential linearized muffin-tin orbital method. Breathing-mode relaxations were included. Formation entropies resulting from the change to the phonon spectrum upon formation of a defect were calculated using a valence force field model for the elastic contributions and a rigid ion model for the Coulomb contributions to the dynamical matrix. The energies and entropies were incorporated into mass-action equations to deduce the relative defect concentration. In agreement with experiments, we find mercury vacancies are the dominant defect to accommodate excess tellurium in the lattice when the material is equilibrated in the presence of mercury vapor; we predict that tellurium antisites are dominant in material equilibrated with tellurium solid. Of the defects that accommodate excess mercury, we find that mercury antisites may be more prevalent than previously thought. Extensions to the alloy are also discussed.

PACS numbers: 68.35.Md, 68.55.Bd

I. INTRODUCTION

Defects are a well known source of problems in HgCdTe and a number of questions concerning them remain unanswered. For one, although post-growth anneals of as-grown p-type vacancy-doped material are routine, the observed conversion to n-type has never been fully satisfactorily explained, although a residual donor is believed to be the culprit. Also, a number of studies on the diffusion of the constituents in HgCdTe and HgZnTe have been done but the source of the observed double-diffusion profiles is still not understood. These and other problems could be better addressed if there were a more complete analysis of the point defects in this material. Moreover, a comparative analysis of the defects in HgCdTe and HgZnTe would facilitate an assessment of the relative merits of these materials.

With the goal of a better understanding of the point defects in HgCdTe and HgZnTe, we examine the properties of the cation and anion vacancies, antisites and interstitial atoms in HgTe. In a previous paper [1] we presented results using the linearized muffin-tin orbital (LMTO) approach within the atomic spheres approximation (ASA) to calculate the defect formation energies. In this paper, we eliminate the ASA and use a more accurate full-potential (FP-LMTO) approach and include a breathing-mode relaxation about the defect site. Additionally, we calculate the defect formation entropy resulting from the change to the phonon spectrum using a Green's function method with a valence force field and rigid ion model. We conclude this paper by incorporating our energies and entropies into a mass-action model for the defect concentrations.

II. DEFECT FORMATION ENERGIES

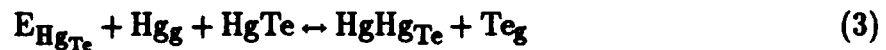
The full-potential linearized muffin-tin orbital method (FP-LMTO) is used to calculate the defect formation energies [2]. These calculations omit the atomic spheres approximation employed in our previous work [1], in which the charge density inside a sphere about each atomic site is assumed spherically symmetric. The FP-LMTO calculations also permit us to calculate the lattice relaxation about the defect site. Defect formation energies are calculated from a difference in total energies of the compound with and without the defect, where the free atom state is used as the reference state, unless otherwise specified. Supercells are used in which a periodic array of defects is constructed. Because we wish to calculate the formation energies in the dilute limit, we use the largest supercell as is computationally feasible. Details of this work, including energy convergence with cluster size, will be presented elsewhere [3].

Two forms of atomic relaxation are used in these calculations. The first is a radial breathing-mode relaxation of the neighbors nearest to the defect site. Second, an overall relaxation of the lattice constant is included to eliminate the pressure on the supercell. This second relaxation will approach zero as the supercell size increases and the dilute limit is approached. We are currently looking at non-breathing-mode relaxations, such as the trigonal and tetragonal distortions which split the degeneracy of the triply degenerate T₂ states [3]. Because the symmetry of the distortion depends on the ionization state of the defect, distortions and charge states must be treated simultaneously.

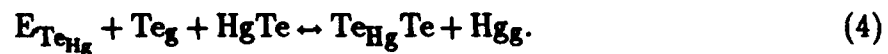
The localized states associated with defects are among the most important and interesting of defect properties, but these properties are difficult to calculate using modern first-principles theories. The local density approximation (LDA), on which the LMTO is based, is inadequate to predict the semiconductor band gaps; higher-order corrections to the many-body equations must be included to correct the band-gap problem. Consequently, in the present calculations we have not attempted to identify the location of the localized states in the gap.

The present calculations are performed using the Harris-Foulkes approximation, in which atomic charge densities are superposed and the resulting charge density and potentials are used to solve the LDA equations. This approximation has been found to be in good agreement with self-consistent results for bulk semiconductor and metal properties [2]. Self-consistent calculations were also done to test the validity of the Harris-Foulkes prescription for defect total energies. Agreement within less than 0.2 eV was found for all defects studied.

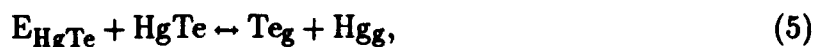
We calculate the defect formation energies for the following reactions:



and



The reference states are chosen so that the number of unit cells on both sides of the reactions are equal and excess atoms are accommodated in their free atomic state. Other reference states of interest can be obtained by considering the additional defect reactions:



and



In Eqs.(1)–(7) A_B corresponds to an A atom or defect on a B site, where V is a vacancy, g the free atom state, and s the elemental solid. Our results for the energy of Eqs. (1)–(5) are given in Table I. The predicted cohesive energy of 4.1 eV is larger than the experimental value of 3.3 eV and is due to underlying errors in the LDA formalism. We are currently considering a number of approaches that would allow us to correct the numbers in Table I for this LDA error. Calculations were done for 8, 16, and 32 atom supercells; for all defects, convergence is found within less than 0.5 eV. We are currently examining larger supercell sizes to better understand the convergence to the dilute limit, which we wish to imitate; results will be presented elsewhere [3].

Nearest-neighbor radial relaxations are also given in Table I. The near-neighbor relaxations are found to be nearly independent of supercell size. Both the Hg and Te antisites are too big to fit into the lattice and result in an outward relaxation of their nearest neighbors. In contrast, the Hg and Te vacancies result in an inward relaxation of the nearest neighbors. These results are in contrast to results for the silicon vacancy [4] where an outward relaxation was found.

III. DEFECT FORMATION ENTROPIES

Defect formation entropies are calculated from the change to the phonon spectrum resulting from the introduction of the defect. A valence force-field model within the rigid hybrid approximation is used to calculate the elastic contributions to the dynamical matrix in the harmonic approximation, and experimental elastic constants are used. For II-VI semiconductors, we also include a Coulomb contribution to the dynamical matrix which results from the ionic nature of the bonding. Unlike the elastic contributions to the dynamical matrix, the Coulomb contributions are long-range in nature and induce a macroscopic electromagnetic field, which results in a screening of the transverse optical phonons. The ionic charge is chosen to yield agreement with experiment for the zone center splitting of the transverse and longitudinal optical phonons. The phonon dispersion

curves calculated from the dynamical matrix are in fair agreement with the experimental dispersion curves. The discrepancies with the experimental curves are largely due to the absence of long-range elastic interactions in the near-neighbor valence force-field model employed.

A Brillouin zone integration is done to calculate the phonon density-of-states for the ideal crystal without a defect. From the density-of-states matrix, $\bar{\rho}$, the Green's function can be calculated from

$$\bar{G}^0(\omega^2) = \int_{\omega_{\min}^2}^{\omega_{\max}^2} \frac{\bar{\rho}(\omega'^2) - \bar{\rho}(\omega^2)}{\omega'^2 - \omega^2 - i\eta} d\omega'^2 + \bar{\rho}(\omega^2) \ln \left(\frac{\omega^2 - \omega_{\min}^2 - i\eta}{\omega^2 - \omega_{\max}^2 - i\eta} \right) \quad (8)$$

where the singularity in the integral has been explicitly removed. The change in the total density-of-states when a defect is introduced into the crystal can be deduced from Dyson's equation to obtain

$$\Delta\rho(\omega^2) = \frac{1}{\pi} \text{Im} \frac{\partial}{\partial \omega^2} (\ln(\det(1 - \bar{G}^0(\omega^2)\bar{V}))) \quad (9)$$

where \bar{V} is the perturbation potential. In the present case, we use a strictly site-diagonal perturbation potential for the mass change due to the introduction of an isolated defect.

The resulting entropy change is given by

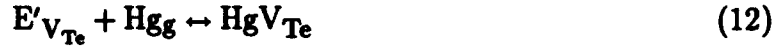
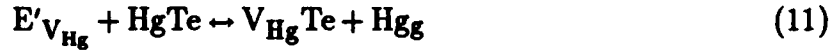
$$S = 2k \int_0^{\omega_{\max}} \Delta\rho(\omega^2) \left(\frac{\hbar\omega}{2kT} \coth\left(\frac{\hbar\omega}{2kT}\right) - \ln\left(2\sinh\left(\frac{\hbar\omega}{2kT}\right)\right) \right) \omega d\omega \quad (10)$$

where k is Boltzmann's constant and \hbar is Planck's constant divided by 2π . Calculated entropy changes for the various defect reactions in Eqs. (1)–(5) are given in Table I. Further details of the calculations, including defects with both on and off-diagonal disorder, are given elsewhere [5].

IV. DEFECT EQUILIBRIUM

In the following analysis we consider only native point defects in bulk HgTe. Although surfaces will play an important role in determining the properties of HgCdTe, in particular for the epitaxial films, we will not discuss those effects here. Furthermore, we consider only the neutral defects in the analysis, although the ionized states are certainly important for many defects, and will affect the defect concentrations. Finally, full equilibrium is assumed, i.e., the material is assumed to have been annealed for a sufficiently long time.

For the analysis of the defects, we will find it convenient to reference all of the defects to pure HgTe and Hg in the vapor state. The above reaction become:



In addition to these reactions, we also include those corresponding to the interstitial mercury and tellurium. These are:



and



The subscript I refers to the interstitial site. Entropies and energies for these two defect reactions are given in Table I. We have not calculated the change to the phonon spectrum from adding an interstitial atom to the solid; we approximate it by equating the entropy of an interstitial atom to that of an atom on its ideal lattice site. The energies for the reactions in Eqs. (15) and (16) have been taken from our previous ASA calculations [6]. Because of the large lattice distortions induced by the interstitial atoms, these calculations are more difficult to perform than for the other defects. Results for interstitial mercury and tellurium using the FP-LMTO will be presented elsewhere [3]. Other defects such as the Schottky and Frenkel type can be obtained by taking sums and differences of the above reactions, Eqs. (11)–(16).

The reaction constants corresponding to the above reactions are:

$$K_{V_{Hg}} = (kT)^{\frac{5}{2}} (2\pi m_{Hg})^{\frac{3}{2}} h^{-3} \exp\left(\frac{S_{V_{Hg}}}{k}\right) \exp\left(\frac{-E'_{V_{Hg}}}{kT}\right) = [V_{Hg}] P_{Hg} \quad (17)$$

$$K_{V_{Te}} = (kT)^{-\frac{5}{2}} (2\pi m_{Hg})^{-\frac{3}{2}} h^3 \exp\left(\frac{S_{V_{Te}} - S_{HgTe}}{k}\right) \exp\left(\frac{-E'_{V_{Te}}}{kT}\right) = [V_{Te}] P_{Hg}^{-1} \quad (18)$$

$$K_{HgTe} = (kT)^{-5} (2\pi m_{Hg})^{-3} h^6 \exp\left(\frac{S_{HgTe} - S_{HgTe}}{k}\right) \exp\left(\frac{-E'_{HgTe}}{kT}\right) = [HgTe] P_{Hg}^{-2} \quad (19)$$

$$K_{TeHg} = (kT)^5 (2\pi m_{Hg})^3 h^{-6} \exp\left(\frac{S_{TeHg} + S_{HgTe}}{k}\right) \exp\left(\frac{-E'_{TeHg}}{kT}\right) = [TeHg] P_{Hg}^2 \quad (20)$$

$$K_{HgI} = (kT)^{-\frac{5}{2}} (2\pi m_{Hg})^{-\frac{3}{2}} h^3 \exp\left(\frac{S_{HgI}}{k}\right) \exp\left(\frac{-E'_{HgI}}{kT}\right) = [HgI] P_{Hg}^{-1} \quad (21)$$

and

$$K_{TeI} = (kT)^{\frac{5}{2}} (2\pi m_{Hg})^{\frac{3}{2}} h^{-3} \exp\left(\frac{S_{TeI}}{k}\right) \exp\left(\frac{-E'_{TeI}}{kT}\right) = [TeI] P_{Hg}. \quad (22)$$

Here T is the temperature in Kelvin, P_{Hg} is the mercury pressure in atmospheres, h is Planck's constant, and m_{Hg} is the mass of mercury. The square brackets indicate the defect site fraction.

We will first consider the defects that accommodate excess tellurium: V_{Hg} , $TeHg$, and TeI . Evaluating the above mass-action relations for the solid in equilibrium at $T=250^\circ C$ and $P_{Hg}=1$ atm, we find

$$\frac{[TeHg]}{[V_{Hg}]} \approx 10^{-9} \quad (23)$$

and

$$\frac{[TeI]}{[V_{Hg}]} \approx 10^{-24}. \quad (24)$$

We can see from these equations that the mercury vacancy is the dominant native point defect to accommodate excess tellurium and that the tellurium antisite and interstitial tellurium densities are negligible in comparison for the given anneal conditions. Even at higher temperatures, near $500^\circ C$, the vacancy is the dominant defect and exceeds the tellurium antisite concentrations by $\sim 10^5$.

Next we consider the equilibrium of the material with elemental solid tellurium. This will be of interest if the material contains tellurium precipitates and the region near the precipitate is in equilibrium with this solid tellurium phase. Rewriting the mass-action equations to reference all defects to the tellurium solid, we find at $T=250^\circ C$ and $P_{Hg}=1$ atm that

$$\frac{[TeHg]}{[V_{Hg}]} \approx 10^4 \quad (25)$$

and

$$\frac{[TeI]}{[V_{Hg}]} \approx 10^{-28}. \quad (26)$$

The result in Eq. (25) indicates that the concentration of tellurium antisites will exceed that of mercury vacancies in the vicinity of the tellurium precipitate. This may have interesting consequences for the properties of partially annealed material when tellurium precipitates are present. Perhaps a more relevant situation to consider is material which has been equilibrated at higher temperatures, which is then subsequently annealed to remove the Hg vacancies. For equilibration at 500°C we find

$$\frac{[\text{Te}_{\text{Hg}}]}{[\text{V}_{\text{Hg}}]} \approx 1000. \quad (27)$$

Even at this elevated temperature the tellurium antisite density is expected to exceed the mercury vacancy density near tellurium inclusions. We would expect that the tellurium antisites to be less mobile than mercury vacancies and thus be more difficult to anneal. Therefore, if the material is slow to anneal at the lower annealing temperatures because of low mobility of the tellurium antisite, an atmosphere of a significant antisite density will be present about the tellurium inclusions. The size of these atmospheres will depend on the dynamics of the growth process, since this is a nonequilibrium situation. If such atmospheres of tellurium antisites are present, we can expect that they will result in some interesting properties. For example, if the tellurium antisites are donors, regions of n-type material may be present in the otherwise p-type, vacancy-doped material at high temperatures. If the tellurium antisites are acceptors, regions of excessive p-type carrier concentration will exist in the p-type, vacancy doped, material. If the tellurium antisites were slow to anneal at low temperatures, subjecting the material to a low temperature anneal could leave regions of relatively high n- or p- type character.

Next, we consider the defects that accommodate excess mercury in the lattice. Using the Hg vapor as the reference state, for material equilibrated at $T=250^\circ\text{C}$ and $P_{\text{Hg}}=1$ atm we find

$$\frac{[\text{Hg}_{\text{Te}}]}{[\text{V}_{\text{Te}}]} \approx 10^{17} \quad (28)$$

and

$$\frac{[\text{Hg}_{\text{I}}]}{[\text{V}_{\text{Te}}]} \approx 10^8. \quad (29)$$

We thus conclude that the the mercury antisite is the dominant defect that accommodates excess mercury in the lattice. Consequently we find that

$$\frac{[\text{Hg}_{\text{Te}}]}{[\text{Hg}_{\text{I}}]} \approx 10^9. \quad (30)$$

Diffusion measurements in HgCdTe indicate that interstitial mercury is the dominant diffusing species at high mercury pressures[7]. Even though we predict them to be a majority species, the mercury antisites may not contribute to diffusion because they are expected to be far less mobile than interstitial mercury and their diffusion will involve tellurium vacancies or interstitial tellurium, both of which are energetically costly to create. Additionally, although the mercury antisites are predicted to be the dominant equilibrium species, this equilibrium may not be realized because the equilibration time for antisites may be substantially longer than for interstitial atoms.

The above results are for the compound HgTe. The question remains as to the extension of these results to the alloys HgZnTe and HgCdTe for the mercury-rich regimes. In a previous paper [8], using a tight-binding model we showed that the vacancy formation energies are sensitive to the alloy environment, in particular for vacancies of the common constituent, i.e., tellurium vacancies in HgCdTe and HgZnTe. Even for vacancies of the substituted species, for which the alloy environment lies in the second neighbor shell and beyond, we found the vacancy formation energies to vary in going from one environment to another, by several tenths of an electron volt. We previously found that the Hg vacancy formation energy is smaller in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ than $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$ for their respective x corresponding to a band gap of 0.1 eV, and indicating lower vacancy concentrations in HgZnTe. For comparisons of the relative defect densities done above, one needs the variation of the antisite and interstitial atom energies, which are also expected to depend on the alloy environment.

Several improvements to the present calculation are under way and will be presented elsewhere [3]. First, we still must confirm the convergence of our calculations by going to even larger supercells. We anticipate that the shifts to the energies will be comparable for all the defects, and therefore conclusions based on ratio of defect densities will be unchanged. Second, as discussed above, a correction for the LDA error would help in improving the present results and allow us to more reliably bench-mark our energies against experimental numbers. Third, we have included only the breathing-mode distortion for the defects. Inclusion of Jahn-Teller distortions can lower the formation energy, but since they are charge-state dependent, we have not done so in the present work. For the neutral defects, we do not expect that this will alter the present conclusions. A combination of the Jahn-Teller distortion and the LDA error is a likely explanation of our discrepancy with the experimental values for the mercury vacancy of 2.24 eV [9]. Finally, the defect equilibrium must be done for the appropriate charge state of the system, while the present analysis was done for the neutral defects only.

ACKNOWLEDGMENTS

The financial support provided for this work through NASA Contract NAS1-18226 and ONR Contract N00014-89-K-0132 is gratefully acknowledged.

REFERENCES

1. M. A. Berding, M. van Schilfgaarde, A. T. Paxton, and A. Sher, *J. Vac. Sci. Technol. A* **8**, 1103 (1990).
2. M. Methfessel and M. van Schilfgaarde (in preparation.)
3. M. A. Berding, M. van Schilfgaarde, and A. Sher (in preparation.)
4. G.A. Baraff, E.O. Kane, and M. Schluter, *Phys. Rev. B* **21**, 5662 (1980).
5. M. A. Berding (in preparation.)
6. M. A. Berding, A. Sher, and A.-B. Chen, *Mat. Res. Soc. Symp. Proc.* Vol. 216, 3 (1991).
7. M.-F. S. Tang and D. A. Stevenson, *J. Vac. Sci. Technol.* **7**, 544 (1989).
8. M. A. Berding, A. Sher, and A.-B. Chen, *J. Appl. Phys.* **68**, 5064 (1990).
9. H. R. Vydyanath, *J. Electrochem. Soc.* **128**, 2609 (1981).

Table I. Defect formation energies for the defect reactions indicated in text.

Energies shown in parentheses are from ASA calculations, published by Berding et al. [6]; an average of the anion and cation tetrahedral interstitial sites has been used here. Defect entropy changes include the phonon contribution only. Radial relaxations are the change in the distance between the center of the defect site and its near neighbors.

	Formation Energy (eV)		Near-Neighbor Radial Relaxation (Å)	S (10^{-5} eV/K)	
	E	E'		250 C	500 C
V _{Hg}	2.6	2.6	-0.14	-66.3	-76.4
V _{Te}	5.8	1.7	-0.17	-60.4	-70.4
Hg _{Te}	3.4	-0.7	0.06	5.5	5.5
Te _{Hg}	-0.1	4.0	0.20	-5.8	-5.8
Hg _{Te}	4.1	-135.5	-155.4
Hg _I	...	(0.88)	...	~ 72	~ 82
Te _I	...	(4.96)	...	~72	~82

Appendix C

**FORMATION ENERGIES, BOND LENGTHS, AND BULK MODULI OF ORDERED
SEMICONDUCTOR ALLOYS FROM TIGHT-BINDING CALCULATIONS**

Formation energies, bond lengths, and bulk moduli of ordered semiconductor alloys from tight-binding calculations

Chin-Yu Yeh and A.-B. Chen

Department of Physics, Auburn University, Auburn, Alabama 36849

A. Sher

SRI International, Menlo Park, California 94025

(Received 20 September 1990)

A tight-binding model is used to calculate the formation energies, bond lengths, and bulk moduli of a number of ordered III-V and II-VI semiconductor alloys. The parameters in the model are adjusted so that the bond lengths, cohesive energies, bulk moduli, and shear elastic constants for the constituent compounds are described correctly. The model is then applied to alloys without further adjustment. Based on the calculated excess energies, we conclude that none of the ordered alloys found experimentally is in its stable bulk equilibrium state at the growth temperatures. Although the alloy excess energies can be negative, if the reference constituent compounds are constrained to match the substrate lattice used in epitaxial growth, their magnitudes are not large enough to account for the observed ordering. A possible explanation of the observed states in terms of a barrier that prevents the metastable ordered alloy from decomposing into separate phases is presented. However, this explanation only applies to alloys with lattice-mismatched constituents. Detailed results on the bond lengths and bulk moduli are also discussed.

I. INTRODUCTION

The bulk semiconductor alloys $A_{1-x}B_xC$ were long thought to be ideal pseudobinary compounds, in which the C atoms sit in a fcc sublattice while the substituting atoms A and B randomly occupy the sites of the other fcc sublattice. However, several recent findings in the last few years revealed quite a different picture. First, extended x-ray-absorption fine-structure (EXAFS) experiments^{1,2} clearly showed a bimodal distribution for the nearest-neighbor bond lengths in these alloys, which implies that the equilibrium atomic positions are not the virtual-crystal sites of the zinc-blende crystal. Recent experiments^{3,4} and theories^{5,6} also indicated that the arrangement of the alloying atoms in these systems is not completely random. Most surprising of all, however, has been the discovery of long-range ordering (LRO) in these alloys grown epitaxially.⁷⁻¹⁹ Essentially, all the III-V alloys grown by molecular-beam epitaxy (MBE) or metal organic chemical vapor deposition (MOCVD) under some special growth conditions are found to be ordered. While a great majority of these ordered alloys form the ABC_2 compounds in one or more of the following three crystal structures: the CuPt (CP), CuAu I (CA), and chalcopyrite (CH), a few alloys are ordered in the form of A_3BC_4 with the famatinite or luzonite structure. These crystal structures have been well described by Wei *et al.*²⁰ Table I is a partial list of the LRO alloys that have been grown, together with the growth conditions and ordered structure found. For later emphasis, we note that the substrate temperatures for the ordering to occur

range from 400°C to 800°C, and the ordering directions are not necessarily the same as the growth direction.

Finding LRO is surprising, because it is at variance with the well-established conventional picture for the bulk semiconductor alloys having simple phase diagrams with miscibility gaps²¹⁻²⁴ driven by strain energy. The question is, are these ordered alloys thermodynamic equilibrium states? This question can be answered if accurate values for the alloy excess energies can be determined. To be specific, we shall only consider three important structures, CP, CA, and CH, for the ABC_2 alloys, and define an excess energy ΔE as

$$\Delta E = E(ABC_2) - [E(AC) + E(BC)], \quad (1)$$

where $E(ABC_2)$ is the energy per molecule, or per four atoms, in the alloy ABC_2 , and $E(AC)$ and $E(BC)$ are the energies per pair of atoms in the AC and BC zinc-blende compounds, respectively. If these ordered alloys are in their thermodynamically stable states at the growth temperature, ΔE has to be negative and must have a magnitude considerably larger than 200 meV on the present scale.²⁵ This would be in contradiction with the positive values of ΔE previously reported for the bulk semiconductor alloys.^{21-24,26} However, the ordered semiconductor alloys found from the epitaxial growth may be in a constrained equilibrium state, where the constraint is imposed by the substrate strain. We need to know the energetics of the various states involved besides the bulk access energy before we can start understanding the ordering and stability of these alloys. The calculation of some of these energies along with an accuracy analysis

TABLE I. A list of the ordered III-V semiconductor alloys identified experimentally.

Alloys	Structure	Growth method	Substrate	Temperature (°C)	Reference
AlGaAs ₂	CA	MOCVD and MBE	GaAs(100) and GaAs(110)	600–800	Ref. 7
AlInP ₂	CP	MOCVD	GaAs(001)	650–700	Ref. 13
AlInAs ₂	CP	MOCVD	InP(001)	600	Ref. 9
GaInP ₂	CP	MOCVD	GaAs(001)	650	Ref. 14
	CP	MOCVD	GaAs(001)	640	Ref. 15
	CP	MOCVD	GaAs(001)	650–700	Ref. 13
	CP	MOCVD	GaAs(001)	600–630	Ref. 16
	CP	MOCVD	GaAs(001)	600–700	Ref. 17
GaInAs ₂	famatinite	LPE	InP(110)	630	Ref. 8
	luzonite	MBE	InP(001)	400	Ref. 10
	CA	MBE	InP(110)	500	Ref. 11
	CP	VLE	InP(001)	650–660	Ref. 12
Ga ₂ AsSb	CA	MOCVD	InP(100)	550–680	Ref. 18
	CH	MOCVD	InP(100)	600	Ref. 18
	CP	MBE		540	Ref. 19

is the main purpose of this paper.

The excess energies for a number of ordered semiconductor alloys have been calculated from local-density-functional (LDF) theory using various band-structure methods.^{6,20,27–32} Although the LDF error for the cohesive energy of a III-V compound is typically several tenths of one eV per pair of atoms, it is generally believed that the excess energy based on the same technique is accurate to several meV, because the errors in LDF cancel in Eq. (1). However, the ordered alloys that we are considering are open structures containing several atoms per unit cell. The atomic positions in these alloys are usually distorted away from the regular zinc-blende sites. Only the most sophisticated band-structure theories which are capable of treating shear distortion, such as the full-potential linear combination of muffin-tin orbitals (FP-LMTO),³³ full potential augmented plane waves (FP-APW),³⁴ and the fully converged plane-wave pseudopotential method,³⁵ can be expected to yield precise results within LDF. Even with present-day computers, it is still too expensive to use these methods to perform calculations over a wide range of semiconductor alloys. On the other hand, although the valence-force-field (VFF) model^{36,37} is simple and is effective in treating the strain energy, it cannot account for the chemical energy.²⁶ These considerations have motivated us to use the empirical tight-binding (ETB) method. The ETB not only can treat both the strain and chemical energies but also allows for precise and systematic computations. To eliminate the propagation of errors from constituent compounds to alloys, the parameters in ETB are adjusted to produce the experimental values for the cohesive energy, bond length d , bulk modulus B , and the shear elastic constant $C_{11} - C_{12}$ for each constituent crystal. These parameters are then used in the alloy calculation without further ad-

justment. Our method thus corresponds to an interpolation scheme for the alloys between the constituent compounds. Such an approach is particularly appropriate for the present study, because the alloys and the constituent crystals have very similar structures and local bonding.

The rest of this paper is arranged as follows: Section II describes the ETB model, the way the parameters are determined, and the results for the structural properties of the constituent crystals. Section III briefly describes the structural parameters and the energy-minimization procedure for the three alloy structures in both their bulk equilibrium states and in states constrained to match substrates. The calculated excess energies are summarized in Sec. IV and are compared with those from LDF and VFF. To provide more detailed structural information, the calculated bond lengths and bulk moduli of the alloys are also presented. The final section, Sec. V, contains a summary and discussion.

II. TIGHT-BINDING MODEL

The tight-binding (TB) model that we are going to use is very similar to that used by Chadi³⁸ and Harrison.^{39,40} The total energy of a semiconductor crystal is assumed to be the sum of the electron energies $\epsilon_v(\mathbf{k})$ in the valence bands and the pair repulsive energies u_{ij} between the nearest-neighbor atoms:

$$E_T = E_{BS} + U_r = \sum_v \sum_k \epsilon_v(\mathbf{k}) + \sum_{i>j} \sum u_{ij}. \quad (2)$$

Furthermore, the band energies are calculated using a minimum-basis TB Hamiltonian which includes one s and three p orbitals per atom. The interaction parameters needed from the Hamiltonian are the term values ϵ ,

and ϵ_p of each atom and the nearest-neighbor two-center interactions $V_{ss\sigma}$, $V_{sp\sigma}$, $V_{pp\sigma}$, and $V_{pp\pi}$. For a given crystal structure, the Hamiltonian $H(\mathbf{k})$ associated with a given wave vector \mathbf{k} within the first Brillouin zone has a dimension of $4m$, where m is the number of atoms per unit cell. The diagonal part of $H(\mathbf{k})$ consists of the term values, and the off-diagonal elements are computed as

$$H_{\gamma\gamma'}(\mathbf{k}) = \sum_{\mathbf{d}} e^{i\mathbf{k}\cdot\mathbf{d}} h_{\gamma\gamma'}(\mathbf{d}_{\gamma\gamma'}), \quad (3)$$

where the sum runs over the first-neighbor bond displacements $\mathbf{d}_{\gamma\gamma'}$ that point from the orbitals denoted γ to γ' . The γ stands for the s , p_x , p_y , or p_z orbital of a particular atom in a given unit cell. The matrix elements $h_{\gamma\gamma'}$ are related to the two-center interactions by the Slater-Koster relations,⁴¹

$$h_{ss} = V_{ss\sigma}, \quad (4)$$

$$h_{sx} = \alpha_1 V_{sp\sigma}, \quad (5)$$

$$h_{xx} = \alpha_1^2 V_{pp\sigma} + (1 - \alpha_1^2) V_{pp\pi}, \quad (6)$$

$$h_{xy} = \alpha_1 \alpha_2 (V_{pp\sigma} - V_{pp\pi}), \quad (7)$$

where $\alpha_i = x_i/d$ are the direction cosines of \mathbf{d} , while all the V 's depend only on the length d .

The first task is to determine the forms for the interactions $V_{aa'}$ and the repulsive energy u from the constituent compounds. Since the strain energy plays a very important role in the alloy formation energy, we shall make sure that our model produces the correct elastic constants. To keep the model close to Harrison's^{39,40} original form, but to free it from his rigid $1/d^2$ and $1/d^4$ scaling rules for $V_{aa'}$ and u , respectively, we assume the following forms:

$$V_{aa'}(d) = V_{aa'}^{(0)}(d_0/d)^n \quad (8)$$

and

$$u(d) = u_0(d_0/d)^m, \quad (9)$$

where the superscript and subscript 0 indicate the values evaluated at the equilibrium bond length d_0 . The values of $V_{aa'}^{(0)}$ are taken to be Harrison's universal forms scaled by a factor f ,

$$V_{aa'}^{(0)} = f V_{aa'}^{\text{Harrison}}. \quad (10)$$

Note that Harrison's⁴⁰ universal two-center interactions take the form

$$V_{aa'}^{\text{Harrison}} = \eta_{aa'} \hbar^2 / (m d_0^2),$$

where m is the free-electron mass, \hbar is Planck's constant, and the η 's take the following values: $\eta_{ss\sigma} = -1.32$, $\eta_{sp\sigma} = 1.42$, $\eta_{pp\sigma} = 2.22$, and $\eta_{pp\pi} = -0.63$.

Thus there are four adjustable parameters for each system: the scaling parameter f , the powers n and m , and the value u_0 . These parameters are determined by requiring that the model produce the correct experimental values for the bond energy E_{bond} , d_0 , B , and the shear elastic constant $C_{11} - C_{12}$. Since $C_{11} - C_{12}$ is governed only by $V_{aa'}^{(0)}$ in the present model, it alone determines the scaling factor f . Then the bond energy E_{bond} is used to determine u_0 . The requirement that the first derivative of E_T be zero at d_0 then determines the ratio of the powers n/m , which couples with the equation for the bulk modulus to yield the values for n and m . One can then use these sets of parameters to calculate other quantities that are not employed in the fitting, e.g., another shear elastic constant C_{44} , the Kleinman internal-displacement parameter⁴² ζ , and the optical-phonon frequencies ω at the zone center, to check the validity of the model.

In the actual calculations we used the term values tabulated by Chen and Sher.⁴³ Table II lists the experimen-

TABLE II. Values of bond length d , bond energy E_{bond} , bulk modulus B , and shear coefficient $C = C_{11} - C_{12}$ used to determine the parameters in Table III. Also listed are the experimental values of C_{44} and the TO-phonon mode ω at Γ to be compared with the calculations. All the elastic constants are in units of 10^{11} dyn/cm², d in Å, E_{bond} in eV, and ω_{TO} is given in terms of wave numbers in cm⁻¹. The sources of these values are discussed in Ref. 47. Also listed are the force constants α and β (N/m) for the valence-force-field model in Eq. (13).

	d	E_{bond}	B	C	C_{44}	ω	α	β
AIP	2.367	-2.13	8.600	6.900	6.150	440	43.867	9.429
AlAs	2.451	-1.89	7.727	7.160	5.420	361	40.360	10.132
AlSb	2.656	-1.76	5.817	4.428	4.076	366	33.417	6.790
GaP	2.360	-1.78	9.143	7.870	7.143	367	46.257	10.723
GaAs	2.448	-1.63	7.690	6.630	6.040	269	40.351	9.371
GaSb	2.640	-1.48	5.792	4.946	4.440	231	32.800	7.539
InP	2.541	-1.74	7.247	4.460	4.600	304	40.346	6.543
InAs	2.622	-1.55	5.794	3.803	3.959	219	33.165	5.757
InSb	2.805	-1.40	4.831	3.130	3.132	185	29.605	5.069
ZnTe	2.637	-1.20	5.090	3.060	3.120	177	29.445	4.659
CdTe	2.806	-1.10	4.210	1.680	2.040	141	26.374	2.722
HgTe	2.798	-0.81	4.759	1.817	2.259	116	29.773	2.935

tal values⁴⁴⁻⁴⁷ for the d_0 , E_{bond} , B , and the $C_{11}-C_{12}$ used to fit the parameters, and the values of C_{44} and transverse phonon frequency ω used for the consistency check.

The elastic constants can be calculated directly from perturbation theory. First the Hamiltonian $H(\mathbf{k})$ is expanded in powers of the infinitesimal strain parameters ϵ up to second order:

$$H = H_0 + H_1\epsilon + \frac{1}{2}H_2\epsilon^2, \quad (11)$$

where H_0 is the strain-free Hamiltonian and H_1 and H_2 are, respectively, the first and second derivatives with respect to ϵ evaluated at $\epsilon=0$. The band-energy contribution to the strain coefficient then comes from the second derivative of E_{BS} with respect to ϵ , denoted by

$$\frac{\partial^2 E_{\text{BS}}}{\partial \epsilon^2} = \sum_v \sum_k \langle v, \mathbf{k} | H_2 | v, \mathbf{k} \rangle + 2 \sum_v \sum_c \sum_k \frac{|\langle v, \mathbf{k} | H_1 | c, \mathbf{k} \rangle|^2}{\epsilon_v(\mathbf{k}) - \epsilon_c(\mathbf{k})}, \quad (12)$$

where $\epsilon_c(\mathbf{k})$ and $|c, \mathbf{k}\rangle$ are, respectively, the eigenenergies and eigenvectors of H_0 for the conduction bands. Similarly, v, \mathbf{k} indexes the valence bands. Note that the intervalence-band contributions in the second-order perturbation sum cancel exactly and so they are not needed in Eq. (12). The matrix elements of H_1 and H_2 needed here can be expressed from Eqs. (3)-(7) in terms of the first and the second strain derivatives of the two-center interactions $V_{aa'}$ and the direction cosines α_i .

Table III shows the results for f , n , m , and u_0 obtained for the constituent compounds, and the corresponding values of C_{44} , ζ , and ω calculated as a consistency check. The scaling factor f ranges from 1.1 to 1.5 and tends to decrease with an increase in polarity. In the power-law dependence $V_{aa'} \propto (d_0/d)^n$, n ranges from 3.3 to 4.2,

which is consistently larger than the $n=2$ used in Harrison's universal TB parameters. For the repulsive pair energy $u = u_0(d_0/d)^m$, the power m ranges from 5.4 to 7.1. The ratio m/n falls in the range from 1.5 to 1.9, which is smaller than the $m/n=2$ used by Harrison.³⁹ The calculated values of C_{44} for most systems agree with experiment to 10% or better. The calculated TO-phonon modes at Γ for most III-V systems also agree with experiments to 10% or better. The discrepancies for the II-VI systems are larger (about 15%). Reliable results for ζ from both experiments and first-principles calculations are only available for a limited number of semiconductors. The calculated ζ , C_{44} , and ω from the TB model agrees very well with those results, as shown in Table IV.

The results in Tables III and IV are based on a particular set of term values and TB parameters. It is useful to examine how the predictions are influenced by these parameters and the fitting procedure. Table V shows a result based on Chadi's procedure^{38,48} in which the TB matrix elements $V_{aa'}$ are scaled as $1/d^2$, and the repulsive pair energy is taken to be

$$u = u_0 + u_1(d - d_0) + u_2(d - d_0)^2.$$

The parameter u_0 is set to produce the correct bond energy, u_1 is determined by requiring the correct equilibrium bond length, and u_2 is fixed by the bulk modulus. Two sets of TB parameters are tabulated for each system: One of them is the set used by Chadi³⁸ and the other (labeled present work) is the set obtained by multiplying Harrison's $V_{aa'}$ by the scaling factor f listed in Table III. For convenient comparison, the zero of the term values is set equal to the anion s energy. Despite considerable differences in these two sets of TB parameters, the results of the predictions from the two sets are very similar and also very similar to those predicted from the other procedure given in Table IV. The only noticeable difference

TABLE III. The results for the parameters f , n , m , and u_0 obtained from the fitting of the bond energy, bond length, bulk modulus, and shear coefficient $C_{11}-C_{12}$ of Table II using the full band-structure calculation. Also listed are the calculated C_{44} , internal-displacement parameter ζ , and the TO-phonon mode ω at Γ . All the elastic constants are in units of 10^{11} dyn/cm², u_0 is in eV, and ω are given in terms of wave numbers in cm⁻¹.

	f	n	m	u_0	C_{44}	ζ	ω
AlP	1.294	3.530	5.598	6.435	5.827	0.516	447
AlAs	1.464	3.524	5.430	7.089	5.598	0.459	384
AlSb	1.337	3.268	5.668	4.838	3.944	0.564	354
GaP	1.395	3.705	5.683	7.285	6.857	0.501	382
GaAs	1.397	3.633	5.716	6.530	5.791	0.500	292
GaSb	1.431	3.471	5.717	5.519	4.515	0.536	256
InP	1.323	4.240	6.633	5.603	4.260	0.584	304
InAs	1.300	3.997	6.427	4.962	3.564	0.552	220
InSb	1.353	3.773	6.399	4.350	3.092	0.602	200
ZnTe	1.284	3.306	5.828	4.285	2.813	0.590	205
CdTe	1.171	3.656	6.761	3.092	1.701	0.694	156
HgTe	1.173	3.760	7.074	3.080	2.040	0.716	152

between the predictions in Table IV and Table V is that in the latter the phonon frequencies are slightly larger and C_{44} are slightly smaller. These comparisons show that the model with parameters given in Table II not only contains the correct structure properties inserted through the fitting procedure, but also predicts other mechanical properties with reasonable accuracy. This model should serve as a good basis for alloy calculations.

III. ALLOY CALCULATION

A. CuAu I structure (CA)

The ABC_2 semiconductor alloy in the CA structure forms a layer structure $ACBCACBC \dots$ along the $\langle 001 \rangle$ direction. The basic lattice vectors can be chosen as $\mathbf{a}_1 = (\frac{1}{2}, \frac{1}{2}, 0)a$, $\mathbf{a}_2 = (\frac{1}{2}, -\frac{1}{2}, 0)a$, and $\mathbf{a}_3 = (0, 0, \beta)a$, where β is the c/a ratio with a and c being the lattice parameters. Note that the ideal β value for a zinc-blende structure is 1. There are four atoms per unit cell: one A atom at $(0, 0, 0)$, one B atom at $(\frac{1}{2}, 0, \beta/2)a$, and two C atoms at $(1, 1, \beta + \delta)a/4$ and $(3, 1, 3\beta - \delta)a/4$, where δ is

TABLE IV. Comparison between calculated and experimental lattice constant a , elastic constants B , $C_{11} - C_{12}$ and C_{44} , Kleinman (Ref. 42) internal-displacement parameter ζ , and the TO-phonon frequency ω in wave numbers cm^{-1} . Also listed are $C_{44}^{(0)}$ which correspond to the value without the internal distortion. The FP-LMTO and PP-PW are the *ab initio* theories, and TB is our tight-binding method discussed in the text. All elastic constants are in units of 10^{11} dyn/cm². Experimental values are those listed in Table II.

		Expt.	FP-LMTO ^a	PP-PW ^b	TB
Si	a	5.431	5.41	5.45	5.431
	B	9.923	9.9	9.3	9.923
	$C_{11} - C_{12}$	10.274	10.2	9.8	10.274
	C_{44}	8.036	8.3	8.5	8.013
	$C_{44}^{(0)}$			11.1	11.30
	ζ	0.54 ^c	0.51	0.53	0.51
	ω	523	518	521	572
Ge	a	5.65		5.59	5.65
	B	7.653		7.2	7.653
	$C_{11} - C_{12}$	8.189		8.5	8.189
	C_{44}	6.816		6.3	6.84
	$C_{44}^{(0)}$			7.7	9.46
	ζ			0.44	0.49
	ω	303		302	342
GaAs	a	5.642		5.55	5.642
	B	7.69		7.3	7.69
	$C_{11} - C_{12}$	6.63		7.0	6.63
	C_{44}	6.04		6.2	5.791
	$C_{44}^{(0)}$			7.5	7.83
	ζ			0.48	0.50
	ω	273		268	292

^aReference 33.

^bReference 49.

^cReference 50.

the internal distortion parameter for the C atoms. There are only two different nearest-neighbor bond lengths in the alloy:

$$d_{AC} = (a/4)[2 + (\beta + \delta)^2]^{1/2}$$

and

$$d_{BC} = (a/4)[2 + (\beta - \delta)^2]^{1/2}.$$

For a set of values for a , β , and δ , the total energy is calculated following the general description in Sec. II. The interactions $V_{\alpha\alpha'}$ and the repulsive energy are scaled by the bond lengths according to Eqs. (8) and (9), respectively. The Hamiltonian $H(\mathbf{k})$ is now a 16×16 matrix. The total energy is then minimized by varying the three parameters a , β , and δ . If the alloy is constrained to match a $\langle 001 \rangle$ substrate, the lattice constant a is fixed by the substrate, and the total energy is minimized with respect to β and δ .

B. Chalcopyrite structure (CH)

The ABC_2 semiconductor alloy in the CH structure forms a superlayer structure $ACACBCBCACACBCBC \dots$ along the $\langle 012 \rangle$ direction. The basic lattice vectors can be chosen as $\mathbf{a}_1 = (1, 1, -2\beta)a/2$, $\mathbf{a}_2 = (-1, 1, 2\beta)a/2$, and $\mathbf{a}_3 = (1, -1, 2\beta)a/2$, where β again is the c/a ratio, with an ideal value of 1. There are now eight atoms per unit cell: two A atoms at $(0, 0, 0)$ and $(0, 1, \beta)a/2$, two B atoms at $(1, 0, \beta)a/2$ and $(1, 1, 0)a/2$, and four C atoms at $(1 + \delta, 1, \beta)a/4$, $(1, 3 + \delta, 3\beta)a/4$, $(3, 1 - \delta, 3\beta)a/4$, and $(3 - \delta, 3, \beta)a/4$, where δ is the internal distortion parameter for the C atoms. Again there are also only two different nearest-neighbor bond lengths in the alloy:

$$d_{AC} = (a/4)[1 + (1 + \delta)^2 + \beta^2]^{1/2}$$

and

$$d_{BC} = (a/4)[1 + (1 - \delta)^2 + \beta^2]^{1/2}.$$

A similar procedure can now be carried out to minimize the total energy with respect to the three parameters a , β , and δ . However, the Hamiltonian $H(\mathbf{k})$ is now 32×32 . For the case in which the lattice is matched to the $\langle 001 \rangle$ substrate, we again are left with two parameters β and δ to vary for the energy minimization.

C. CuPt structure (CP)

In the CP structure, the alloy forms a $\langle 111 \rangle$ superlattice $ACBCACBC \dots$. Because of a lack of reflection symmetry about any of these planes, the B layer need not be located exactly at the middle position between the two successive A layers. Also the distance between two closest atoms from two different A layers may not need to correlate with that between two A atoms on the same plane. Thus there are a total of five independent parameters required to describe the crystal structure: the lateral lattice constant a for the layers, the spacing D between two successive A layers, and the three spacing parameters for the three layers (one B and two C) inside D .

TABLE V. Comparison between the two different sets of TB parameters described in the text; the resultant coefficients u_0 , u_1 , and u_2 of u , and the predicted elastic constants, Kleinman internal-displacement parameters ζ , and phonon frequency ω .

		ϵ_s^A	ϵ_p^A	ϵ_s^C	ϵ_p^C	$V_{ss\sigma}$	$V_{sp\sigma}$	$V_{pp\sigma}$	$V_{pp\pi}$	
Si	Chadi	0.0	7.20	0.0	7.20	-2.03	2.55	4.55	-1.09	
	Present work	0.0	6.88	0.0	6.88	-2.41	2.59	4.05	-1.15	
		u_0	u_1	u_2	C	C_{44}	C_{44}^0	ζ	ω	
Si	Chadi	7.29	-9.98	23.90	10.66	7.89	11.38	0.49	620	
	Present work	6.93	-9.70	23.42	10.27	7.83	11.39	0.51	592	
		ϵ_s^A	ϵ_p^A	ϵ_s^C	ϵ_p^C	$V_{ss\sigma}$	$V_{sp\sigma}^{AC}$	$V_{sp\sigma}^{CA}$	$V_{pp\sigma}$	$V_{pp\pi}$
GaAs	Chadi	0.0	9.64	5.12	11.56	-1.70	2.40	1.90	3.44	-0.89
	Present work	0.0	10.09	6.79	14.12	-2.34	2.52	2.52	3.94	-1.12
		u_0	u_1	u_2	C	C_{44}	C_{44}^0	ζ	ω	
GaAs	Chadi	5.12	-7.12	18.22	6.36	5.60	8.77	0.54	339	
	Present work	6.53	-8.39	19.90	6.63	5.70	8.53	0.54	322	

There are, however, only four atoms per unit cell, each coming from a different layer, so the $H(\mathbf{k})$ matrix is 16×16 . There are also four different bond lengths in the crystal, two for the AC and two for the BC bonds. If the alloy lattice is matched to a (111) substrate, the lateral lattice parameter is fixed, but we still have four parameters to vary for energy minimization. However, if the alloy is matched to a (001) substrate, we shall assume that all the alloying atoms A and B are locked into the fcc lattice points of the substrate, and we are left with only two free parameters which describe the relaxation of the two nonequivalent C layers.

It is useful to comment on the Brillouin-zone (BZ) integration needed for the calculation of the electronic energy. Since our calculation involves relatively small and easily handled matrices, we are able to sample over a large number of \mathbf{k} points. We found that a uniform grid of 1000 \mathbf{k} points inside the BZ always guarantees a convergence of the total energy per atom to within 10^{-3} meV and the elastic constants to an accuracy in the third digit. We found that total energies calculated using two special \mathbf{k} points⁵¹ are about 10 meV per atom pair higher than the converged value. However, when extending the two-special- \mathbf{k} -point method to the CA and CH structures,⁵² most of the errors of the alloy and pure compound cancel, and we found that the final errors in the excess energies are only about 0.5 meV per atom pair.

Before presenting the results, we briefly describe the Keating³⁶ valence-force-field (VFF) model. The VFF only deals with the strain energy. The energy per unit volume in a strained diamond or zinc-blende crystal is given by

$$U = \frac{3}{8d_0^2} \sum_i \alpha_i [\Delta(\mathbf{d}_i \cdot \mathbf{d}_i)]^2 + \frac{3}{8d_0^2} \sum_{i>j} \sum \beta_{ij} [\Delta(\mathbf{d}_i \cdot \mathbf{d}_j)]^2, \quad (13)$$

where in the first term, the bond-stretching energy, the

summation i runs over all the bonds, and in the second term, the bond-angle contribution, the summations include all the pairs of bonds that share common atoms. The d_0 in Eq. (13) is the equilibrium bond length, and $\Delta(\mathbf{d}_i \cdot \mathbf{d}_j)$ is the strain-induced change of the dot product between the two bond vectors which point from the common atom to the nearest-neighbor atoms. For a zinc-blende crystal, there is only one value for the bond-stretching force constants $\alpha_i = \alpha$, and one value for the bond-angle-restoring-force constant $\beta_{ij} = \beta$. Their values are determined²⁶ by fitting the experimental bulk moduli B and shear elastic constants $C_{11} - C_{12}$, and are tabulated in Table II. We extend Eq. (13) to calculate the strain energy in an ABC_2 alloy by treating each bond and each pair of the same bonds in the same way as in the constituent compounds. However, when dealing with the bond-angle term involving two unlike AC and BC bonds, both d_0 and β_{ij} are taken to be the average values. The results from VFF will be compared with the ETB calculations in Sec. IV.

IV. DISCUSSION OF ALLOY RESULTS

A. Excess energies

Before presenting the results for different kinds of excess energies, the readers should be reminded of the relation between these energies and ordering. The Appendix provides a qualitative discussion of this relation. It also describes how different kinds of excess energies presented below may alter the picture of ordering.

The bulk excess energies ΔE calculated from ETB for the ABC_2 alloys in the three structures of CA, CH, and CP are listed in Table VI along with the results from VFF and LDF. First we observe that ETB and VFF give very similar results. Both models produce very small ΔE values for those alloys composed of compounds with nearly equal lattice constants. However, ΔE for AlGaP_2

and HgCdTe_2 derived from ETB are slightly negative. The differences between the ETB and VFF results in these systems are a measure of the small size of the chemical-energy contribution to ΔE in the ETB model. For the lattice-mismatched alloys, all the ΔE values from ETB are positive and are slightly larger than those obtained from VFF except for InAlP_2 . The major reason VFF yields a smaller ΔE is that the VFF used here consistently predicts smaller values of the elastic constant C_{44} than the experimental values, thereby underestimating the strain contribution to ΔE . However, the small chemical energy included in ETB may upset this trend, exemplified by InAlP_2 .

The quantitative comparison between ETB and LDF is mixed. Starting from the lattice-matched alloys GaAlAs_2 , the ΔE from LDF calculated by different groups range from 7.5 to 35 meV for the three structures considered, as compared to nearly zero calculated from ETB. Similar differences between LDF and ETB also occur in another lattice-matched system, HgCdTe_2 . Since the strain energy is nearly zero in these systems, the 10-meV or so difference between the present calculation and the majority of the LDA results represents the discrepancy in the estimate of the chemical energy in ΔE between the two theories. These differences certainly are well within the margins of errors of both ETB and LDF. However, we note the LDA ΔE values for GaAlAs_2 in the CA structure are rather consistent except the 35 meV from Ref. 56. For the lattice-mismatched alloys, ETB agrees very well (within 20 meV) with LDF for the III-V alloys in the CP structure and II-VI alloys in the CH structure. The agreement is also reasonable for all alloys in the CA structure. However, the differences between the two calculations are more substantial for the III-V alloys in the CH structure and II-VI alloys in the CP structure. We note that although the trend $\Delta E_{\text{CH}} < \Delta E_{\text{CA}} < \Delta E_{\text{CP}}$ among the three structures holds for both ETB and LDF for these lattice-mismatched alloys, the ΔE_{CH} values from LDF for the III-V systems are considerably lower than those from ETB, particularly the negative values calculated for InAlP_2 and InAlAs_2 . Despite all these quantitative differences, the qualitative trends of our ETB predictions are rather similar to those from LDF by Wei and co-workers^{6,27} as shown in Fig. 1. The most important conclusion that can be drawn from Table VI is that these ordered bulk alloys are not the thermal equilibrium states at the experimental growth temperatures shown in Table I. For this to happen, the ΔE value has to be less than -200 meV per four atoms,²⁵ as mentioned earlier. This ΔE value is far below all the calculated values and is beyond the uncertainties of our ETB model and the LDF calculations listed above.

While the above excess energies do not support a thermally stable ordered bulk alloy at the growth temperature ($\approx 600^\circ\text{C}$), they may offer different conclusions when applied to the epitaxially grown alloys. The idea was first suggested by Flynn⁵⁷ in connection with epitaxial growth of ordered metal alloys. If the substrate material serves as a reservoir for the epitaxial film, then the alloys grown on the substrate have to be in thermal and mechanical equilibrium with the substrate. Since most

substrates used in the growth are selected to have their lattice constants close to the equilibrium lattice parameters (see Tables I and IX) of the ordered alloys mentioned above, it takes little strain energy for these alloys to match the substrates. On the other hand, if an alloy on the substrate is to decompose into two substrate lattice-mismatched constituent compounds, it takes energy to constrain the separated systems. Only the case of thin layers in which no dislocations form will be treated. Then the reference constituent energies $E(AB)$ and $E(AC)$ in Eq. (1) which must be used to calculate the excess energy for epitaxial growth, referred to as ΔE_{ep} , are the ones with their lattice parameters matched to the substrate. This coherent strain can make ΔE_{ep} negative for an alloy even when its bulk excess energy ΔE is positive. Of course, the actual values of excess energies are substrate specific. However, an overall assessment can be made when the substrate is also an alloy by choosing the substrate lattice constant to be the average value

$$a_s = \bar{a} = (a_{AC}^0 + a_{BC}^0)/2.$$

The major difference between ΔE_{ep} and ΔE then comes from the strain energies of the constituent compounds forced to lattice match the substrate. Since the energies of the ordered alloys at their equilibrium lattice constants a only differ by a small amount (1 or 2 meV) from those at \bar{a} , we shall use the alloy energies already calculated in Table VI to deduce the epitaxial excess energy.

The strain energy of either constituent compound matched to $a_s = \bar{a}$ on the (001) substrate can be estimated from the following simple formula:

$$E_s(AC) = (C_{11} + 2C_{12} - 4C_{12}^2/C_{11})\delta_0^2 a_s^3/4, \quad (14)$$

where $\delta_0 = (a_{AC}^0 - a_{BC}^0)/\bar{a}$ is the percentage lattice-parameter difference between the two constituent compounds. The strain energies used to calculate the values of ΔE_{ep} in Table VII, however, are obtained from ETB, which in fact only produces a small correction (1 to 2

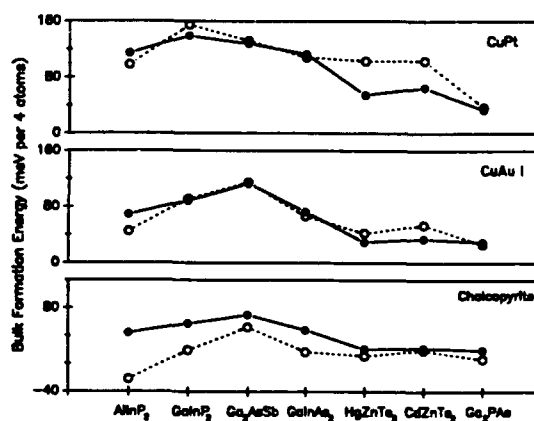


FIG. 1. The excess energies of Eq. (1) for the ordered alloys in three crystal structures calculated from the present TB model (solid circles) and from LDF in Ref. 6 (open circles).

meV at most) to Eq. (14). The relative values of ΔE_{ep} among the three structures on the same substrate, say, (001), should remain the same as those of ΔE , because these two energies only differ in the reference energy. For example, the relation

$$\Delta E_{ep}(\text{CH}) < \Delta E_{ep}(\text{CA}) < \Delta E_{ep}(\text{CP})$$

still holds for alloys composed of lattice-mismatched constituent compounds. On the (001) substrate, the ΔE_{ep} values are essentially all positive in the CP, negative in the CH, and switch between positive and negative values in the CA structure. The CP alloys matched to the (111) substrate have very small ΔE_{ep} with magnitudes smaller than 10 meV. The LDF epitaxial energies calculated by Wei, Ferreira, and Zunger⁶ have the same qualitative trends as those from ETB among the three structures. The quantitative differences in ΔE_{ep} between the two calculations are larger and also more scattered than those in ΔE . This is peculiar, because the difference between ΔE and ΔE_{ep} in both calculations comes from the strain energies of the constituent compounds, which can be reliably estimated from Eq. (14) and are well prescribed by our

ETB model. Despite these discrepancies, a simple conclusion can also be drawn from Table VII. These numbers still cannot account for the ordering observed experimentally, because all the calculated ΔE_{ep} are still well above the -200 -meV value required for the ordering at the growth temperatures.²⁵

In the above consideration, the constituent compounds are constrained to match the lateral lattice of the substrate and are allowed to relax fully in the growth direction. For a (001) substrate with a lattice constant $a_s = \bar{a}$, the c/a ratio for either compound is estimated to be

$$c/a = 1 + \delta_0(C_{12}/C_{11} + 0.5), \quad (15)$$

where δ_0 again is the percentage bond-length difference between the two constituent compounds. This c/a relaxation results in elongation for one compound and shrinkage for the other along the growth direction. If the lattice constants of two constituent crystals have a substantial difference, these opposite relaxations will produce strained grain boundaries between the AC and BC crystals. The reference state energy $E(AC) + E(BC)$ used in

TABLE VI. The bulk excess energies ΔE (in meV per four atoms) from ETB and comparison with the results from local-density-functional (LDF) theory and the valence-force-field (VFF) model.

Alloys	CuAu I			Chalcopyrite			CuPt		
	ETB	LDF	VFF	ETB	LDF	VFF	ETB	LDF	VFF
AlGaP ₂	-2.4		1.5	-2.8		1.0	-2.6		0.2
AlGaAs ₂	0.6	10.8 ^a 11.5 ^c 15.1 ^b 13.5 ^d 35 ^e	0.3	0.6	11.4 ^b 9.8 ^c	0.2	0.8	7.5 ^c	0.4
AlInP ₂	69.0	43.0 ^f	74.0	44.0	-21.0 ^l	47.3	114.2	97.0 ^f	111.1
AlInAs ₂	68.8	35.0 ^f	66.6	45.8	-15.0 ^l	43.7	107.0		97.6
Ga ₂ PAs	30.0	26.6 ^g	23.1	19.4	6.5 ^g	15.0	33.6	37.2 ^g	32.7
Ga ₂ PSb	260.2		207.0	67.4		135.8	290.8		292.6
Ga ₂ AsSb	113.0	129.2 ^a 114.8 ^b 115.0 ^c	91.8	69.8	52.0 ^c	59.9	128.0	132.0 ^c	130.1
GaInP ₂	88.4	115.6 ^a 91.0 ^f 54.4 ^k	83.8	57.2	19.0 ^f	54.3	139.4	155.4 ^g	124.9
GaInAs ₂	73.2	60.1 ^f 85.6 ^a 66.7 ^f	67.3	48.2	16.5 ^g	43.8	113.0	108.5 ^g	99.5
GaInSb ₂	57.4		51.5	37.4		33.2	85.58		76.2
ZnCdTe ₂	34.3	54.2 ^g	33.7	22.4	19.2 ^g	21.1	65.3	103.5 ^g	56.0
ZnHgTe ₂	29.7	42.5 ^g	32.1	21.0	11.4 ^g	19.9	54.9	103.3 ^g	53.5
CdHgTe ₂	-2.3	12.1 ^g	0.61	-2.7	11.3 ^g	0.4	-2.7	9.8 ^g	1.1

^aReference 28.

^bReference 31.

^cReference 27.

^dReference 55.

^eReference 56.

^fReference 53.

^gReference 6.

^hReference 30.

ⁱReference 32.

^jReference 29.

^kReference 54.

^lReference 74.

TABLE VII. The excess energy ΔE_p (in meV per four atoms) for epitaxial alloys calculated from the present TB model and comparison with the LDF results.

Structure substrate (Alloy)	CuAu I (100)		Chalcopyrite (100)		(111)	CuPt (100)	(100)
	TB	LDF	TB	LDF	TB	TB	LDF
AlGaP ₂	-2.9		-3.1		0.5	-3.0	
AlGaAs ₂	0.8		0.7		0.4	0.2	
AlInP ₂	-11.5	-36.0 ^a	-36.3		-0.2	33.8	-18.0 ^a
AlInAs ₂	-4.6		-27.8		-7.4	33.6	
Ga ₂ PAs	5.3	4.1 ^a	-5.3	-21 ^a	0.2	8.9	0.5 ^a
Ga ₂ PSb	40.5		-52.2		7.8	71.1	
Ga ₂ AsSb	15.5	35.0 ^a	-27.6	-28 ^b	-5.6	30.5	52.0 ^b
		17.0 ^c		-45.8 ^a			11.5 ^a
		48.0 ^c					
GaInP ₂	-3.3	-3.0 ^c	-34.5	-106 ^a	-9.4	47.6	3.4 ^a
		12.8 ^c					
GaInAs ₂	0.4	-9.7 ^a	-25.4		-9.6	39.5	4.1 ^a
		10.5 ^d					
		6.7 ^d					
		8.4 ^d					
		29.6 ^c					
GaInSb ₂	1.4		-18.5		-5.0	29.9	
ZnCdTe ₂	-5.6	-5.4 ^a	-17.5	-57.2 ^a	6.6	25.5	0.2 ^a
ZnHgTe ₂	-5.9	-13.5 ^a	-14.7	-64.2 ^a	3.0	19.3	0.9 ^a
CdHgTe ₂	-2.4		-2.8		-3.0	-2.9	

^aReference 6.^bReference 27.^cReference 28.^dReference 29.

the calculation of ΔE_p in Table VII assumed that the decomposed AC and BC phases are macroscopic crystals. The domain-wall energy contributions were neglected because there are few boundaries. However, a realistic path between the completely separated and fully relaxed AC and BC domains is likely to pass through a sequence of intermediate states, including the stage of forming microscopic AC and BC clusters which serve as nucleation centers. As an approximation to this phase space immediately adjacent to the ordered ABC₂ alloy, we have estimated the metastable nucleation energy barrier by assuming that the microscopic AC and BC clusters are lattice matched to the ABC₂ alloy and their energetics can be estimated from the bulk crystals under the same constraint. Because the *c/a* ratios for the ordered alloy are nearly unity (see Table VIII), this epitaxial energy against nucleation of AC and BC clusters is equivalent to using a reference energy $E(AC) + E(BC)$ which disallows the *c/a* relaxation. Consequently, this epitaxial excess energy for the alloy, referred to as ΔE_p^H (the hard model), is lowered further as shown in Table VIII. Note that the values for a number of alloys have already attained magnitudes that could account for the stability of the observed LRO of lattice-mismatched alloys. However, details of the mechanism that causes the LRO to preferentially form in the first place must still be determined.

B. Bond lengths

In addition to the excess energies, our calculations also produce detailed information about the equilibrium structures of the alloys. Table IX lists the *c/a* ratio and

TABLE VIII. The excess energy ΔE_p^H (in meV per four atoms) for epitaxial alloys without (*c/a*)-ratio relaxation.

	CuAu I	Chalcopyrite	CuPt
AlGaP ₂	-3.1	-3.4	-3.3
AlGaAs ₂	0.7	0.6	0.2
AlInP ₂	-192.5	-217.4	-147.3
AlInAs ₂	-149.0	-172.2	-110.8
Ga ₂ PAs	-37.5	-48.1	-33.9
Ga ₂ PSb	-354.4	-447.1	-323.8
Ga ₂ AsSb	-162.2	-205.2	-147.1
GaInP ₂	-200.4	-231.6	-149.5
GaInAs ₂	-149.8	-174.8	-109.9
GaInSb ₂	-113.9	-133.8	-85.3
ZnCdTe ₂	-131.6	-143.5	-100.5
ZnHgTe ₂	-150.3	-159.1	-125.1
CdHgTe ₂	-2.8	-3.2	-3.2

the first-neighbor bond lengths for all the alloys studied. These results are particularly useful if experiments such as EXAFS are carried out to measure the local structures of these LRO alloys. As mentioned earlier, the c/a ratios are nearly unity for all alloys. The notations used in the table are such that the equilibrium bond lengths of the

constituent crystals are in the order $d_{AC}^0 > d_{BC}^0$. The bond lengths in the CA and CH structures are in general bimodal, with $d_{AC} > d_{BC}$. This result is similar to that found in the disordered bulk alloys. There are four different bond lengths in the CP structure: d_{AC}^T , d_{AC}^S , d_{BC}^T , and d_{BC}^S , where superscripts T and S mean the triple

TABLE IX. The equilibrium bond lengths d_{AC} and d_{BC} and the equilibrium average lattice constants a (all in Å). In CP structure, the first values d_{AC} and d_{BC} are for those bonds along the (111) direction, and the second values are for those in the other three directions.

ABC_2		c/a	a	d_{AC}		d_{BC}	
AlGaAs ₂	CA	1.000	5.658	2.450		2.449	
	CH	1.000	5.656	2.450		2.449	
	CP	0.998	5.660	2.451	2.449	2.444	2.451
AlGaP ₂	CA	1.000	5.457	2.365		2.362	
	CH	1.000	5.459	2.366		2.361	
	CP	0.998	5.461	2.371	2.362	2.354	2.366
InGaSb ₂	CA	1.009	6.272	2.781		2.669	
	CH	0.995	6.298	2.788		2.660	
	CP	1.001	6.295	2.788	2.762	2.650	2.699
InAlAs ₂	CA	1.013	5.831	2.597		2.479	
	CH	0.992	5.870	2.604		2.469	
	CP	1.001	5.861	2.599	2.578	2.456	2.512
InGaAs ₂	CA	1.012	5.834	2.597		2.478	
	CH	0.993	5.866	2.605		2.468	
	CP	1.002	5.859	2.602	2.577	2.457	2.510
InAlP ₂	CA	1.007	5.658	2.520		2.395	
	CH	0.996	5.674	2.527		2.384	
	CP	1.001	5.677	2.516	2.504	2.381	2.425
InGaP ₂	CA	1.013	5.640	2.518		2.390	
	CH	0.993	5.672	2.526		2.379	
	CP	1.003	5.665	2.520	2.499	2.368	2.425
AsPGa ₂	CA	1.001	5.549	2.430		2.379	
	CH	0.999	5.552	2.437		2.371	
	CP	0.998	5.556	2.450	2.422	2.354	2.390
SbAsGa ₂	CA	1.003	5.880	2.601		2.499	
	CH	0.998	5.884	2.613		2.482	
	CP	0.995	5.905	2.637	2.585	2.439	2.531
SbPGa ₂	CA	1.008	5.776	2.586		2.435	
	CH	0.994	5.787	2.603		2.407	
	CP	0.993	5.822	2.635	2.567	2.344	2.486
CdHgTe ₂	CA	1.000	6.471	2.805		2.799	
	CH	1.000	6.471	2.806		2.798	
	CP	0.999	6.473	2.817	2.798	2.788	2.805
HgZnTe ₂	CA	1.010	6.256	2.783		2.656	
	CH	0.995	6.286	2.787		2.650	
	CP	1.003	6.277	2.771	2.770	2.659	2.679
CdZnTe ₂	CA	1.016	6.252	2.790		2.657	
	CH	0.991	6.302	2.795		2.650	
	CP	1.007	6.275	2.787	2.771	2.647	2.685

TABLE X. The ratios $\gamma \equiv (d_{\text{all},\gamma} - \bar{d}) / (d_0 - \bar{d})$ for the ordered alloys calculated from the present TB model and comparison with experimental results for the disordered alloys.

ABC ₂	CA ETB		CH ETB		Disordered alloys Experiments	
	γ_{AC}	γ_{BC}	γ_{AC}	γ_{BC}	γ_{AC}	γ_{BC}
InGaSb ₂	0.71	0.64	0.80	0.76	0.89 ^a	0.88 ^a
InGaAs ₂	0.71	0.66	0.80	0.78	0.77 ^a	0.80 ^a
InGaP ₂	0.74	0.67	0.83	0.79	0.80 ^a	0.76 ^a
AsPGa ₂	0.59	0.57	0.75	0.75	0.76 ^a	0.75 ^a
HgZnTe ₂	0.81	0.76	0.86	0.84	0.72 ^b	0.73 ^b

^aReference 58.

^bReference 2.

and single bonds, respectively. However, the alloy lattice constants a are quite close to the mean value \bar{a} of the constituent compounds. A more sensitive measure of the bimodal distribution, following Boyce and Mikkelsen,¹ is to look at the ratio $\gamma_{AC} = (d_{AC} - \bar{d}) / (d_{AC}^0 - \bar{d})$ with \bar{d} being the average bond length, and similarly, γ_{BC} . Table X compares these ratios for the CA and CH structures with the experimental values for the disordered bulk alloys. Note that a value of $\gamma = 1$ corresponds to the totally relaxed case where there is no bond stretching, whereas $\gamma = 0$ corresponds to a rigid virtual crystal where all the atoms are on the zinc-blende crystal sites. In terms of the valence-force model, this ratio is roughly $\gamma = \alpha / (\alpha + 2\beta)$, where α and β [see Eq. (13)] are the average values for the bond-stretching and angular-restoring-force constants of the two constituent compounds. The calculated γ ratios in Table X follow this trend (see the α and β values in Table II). However, the values of γ show that CH structures are more relaxed than the CA structures. This result is consistent with the lower ΔE values in CH than in CA shown in Table VI. The ETB values of γ for GaInAs₂, GaInP₂, and Ga₂AsP in the CH structure are very close to the experimental values^{2,58} for the disordered alloys.

This result, when correlated with the lower excess energy, may suggest that the disordered alloys tend to favor a local configuration of the CH structure. However, the calculated γ values for GaInSb₂ and HgZnTe₂ do not correlate well with the experimental results. As a matter of fact, the trend as a function of the β/α ratio in the experimental results is reversed. It would be interesting to see if these two ordered alloys do have different bond lengths from the disordered states. While the values of γ in this table range from 0.6 to 0.9 for the lattice-mismatched alloys, this trend does not hold for the lattice-matched alloys. For example, Table IX shows that for Hg—Te and Cd—Te bond lengths in HgCdTe₂, both CH and CA structures nearly retain their respective constituent crystal values (i.e., $\gamma \approx 1$), which was also seen in a recent experiment on the bulk alloy.⁵⁹ Finally, the bond lengths for the CP structure are characteristically different from those in CA and CH structures. The single bonds along the ordering direction [111] tend to be close to the constituent values while the triple bonds in the other directions have less relaxations (with γ values around 0.5 or less).

TABLE XI. Bulk moduli (in 10^{11} dyn/cm²) of ordered alloys in three crystal structures calculated from the present TB model and the percentage deviations $\Delta B = (B - B_{av}) / B_{av}$ from the average values B_{av} of the constituent compounds.

	B_{CH}	B_{CA}	B_{CP}	B_{av}	ΔB_{CH} (%)	ΔB_{CA} (%)	ΔB_{CP} (%)
AlGaP ₂	8.858	8.858	8.854	8.8715	-0.150	-0.15	-0.201
AlGaAs ₂	7.695	7.693	7.689	7.7085	-0.181	-0.20	-0.257
AlInP ₂	7.876	7.860	7.774	7.9235	-0.605	-0.08	-1.882
AlInAs ₂	6.705	6.691	6.661	6.7605	-0.828	-1.03	-1.475
Ga ₂ PA ₃	8.328	8.291	8.294	8.4165	-1.046	-1.50	-1.457
Ga ₂ PSb	6.584	6.297	6.188	7.4675	-11.836	-15.68	-17.135
Ga ₂ ASb	6.314	6.198	6.157	6.7410	-6.342	-8.05	-8.662
GaInP ₂	8.007	8.035	8.878	8.1950	-1.437	-1.95	-3.865
GaInAs ₂	6.610	6.579	6.508	6.7420	-1.961	-2.42	-3.474
GaInSb ₂	5.226	5.202	5.156	5.3115	-1.607	-2.07	-2.923
ZnCdTe ₂	4.6105	4.6035	4.3375	4.6500	-0.85	-1.00	-6.72
ZnHgTe ₂	4.8898	4.8872	4.6323	4.9245	-0.71	-0.76	-5.93
CdHgTe ₂	4.4697	4.4721	4.4706	4.4845	-0.33	-0.28	-0.31

C. Bulk moduli

Table XI lists the results for the bulk moduli calculated from ETB. Also listed are the percentage deviations from the mean $\Delta B/\bar{B}$. Note that all ΔB values are negative. The magnitudes of ΔB are small except for Ga_2AsSb and Ga_2PSb which have large differences in both B and d between the constituent compounds. Although the magnitudes of ΔB get larger for systems with larger differences in the bond lengths, the dependence does not seem to be a simple function of the bond-length difference. The uniformly negative ΔB values were also obtained from LDF by Ferreira *et al.*⁶⁰ for the Ga_2AsSb alloys. One reason for the negative values of ΔB , in a very qualitative argument, is due to the fact that the bulk moduli of semiconductors scales inversely as high powers of the lattice constant,⁶¹ and at the same time, the alloy lattice constant is approximated well by the mean value—the Vegard law. This qualitative behavior also comes out of a simple VFF analysis for disordered alloys.⁶²

V. SUMMARY AND CONCLUSION

This work was motivated by our interest in understanding the mechanism of ordering for the semiconductor alloys grown by MBE and MOCVD. We have applied an empirical tight-binding (ETB) model to systematically interpolate the alloy total energies from those of the constituent compounds. Since the strain energy makes a dominant contribution to the excess energy ΔE , particular attention has been given the elastic properties, in addition to the lattice constants and cohesive energies. Our calculated bulk excess energies ΔE are positive for all alloys composed of lattice-mismatched compounds, and nearly zero for the lattice-matched systems. Based on these results, we conclude that all the ordered semiconductor alloys found experimentally are not in their thermodynamic stable states at the experimental growth temperatures. The same conclusion can also be drawn from the most recent LDF calculations^{5,20,27-32,53-56} listed in Table VI. We note that several earlier theories⁶³⁻⁶⁶ that concluded a stable ordering for these alloys have all been revised^{54,67-70} (see also comments in Ref. 71).

Our calculation also generates detailed information about the structures and bulk moduli of these alloys. These results should be checked experimentally.

To further explore the stability of these alloys, we examined the energetics for spontaneous ordering when the grown materials are constrained to match a substrate lattice. Two kinds of epitaxial energies are calculated: One, denoted ΔE_{ep} , corresponds to the situation in which the constituent compounds are allowed to relax fully along the growth direction; and the other, the hard-model ΔE_{ep}^H , does not allow c/a relaxation. Although the sign of ΔE_{ep} can be negative, the magnitudes are too small to account for the observed spontaneous ordering. However, the sign and magnitudes of ΔE_{ep}^H for a number of lattice-mismatched alloys are found to be comparable to the energies needed for ordering at the growth temperatures. Whether or not this is a plausible mechanism

deserves a more detailed investigation.

Even if ΔE_{ep}^H turns out to be a possible mechanism, based on the considerations presented in this paper we are still left without an explanation for the ordering in the lattice-matched alloys such as GaAlAs_2 . We do not believe that any refined bulk calculation will produce an excess energy with large enough magnitude to account for the observed ordering in GaAlAs_2 . On the other hand, epitaxial growth is very surface sensitive. Because of changes in the bonding character at surfaces, e.g., dangling bond, charge transfer, and reconstruction, the surface structural energies behave quite differently from those in the bulk, and do not extrapolate from the bulk, energies.⁷² New mechanisms for spontaneous ordering may emerge from surface energetic considerations. Some hopeful thoughts⁷³ along this line have already been suggested.

ACKNOWLEDGMENTS

The work was supported in part by Office of Naval Research Contract No. N00014-88-C0096 and Air Force Office of Scientific Research Contract No. F4920-88-K-0009. The calculation was done by using the Alabama Supercomputer Network. The authors thank Dr. Alex Zunger for providing useful information.

APPENDIX: EXCESS ENERGY AND ORDERING

The excess energies ΔE , ΔE_{ep} , and ΔE_H considered in the text for the ordered alloys have to be considered along with the free energies of other possible phases including the disordered alloy in order to determine if the ordered phase is stable at a given temperature. A schematic mixing free-energy curve ΔF for a disordered pseudobinary alloy $A_{1-x}B_xC$ as a function of alloy composition x at several different temperatures is shown in Fig. 2. Figure 2(a) is for the case $\Delta E > 0$, and Fig. 2(b) for $\Delta E < 0$. (These curves are similar to Fig. 1 in Ref. 57.) If $\Delta E > 0$, then the stable phase is either the segregated phase or a uniform disordered solution depending on whether the temperature T is lower or greater than the

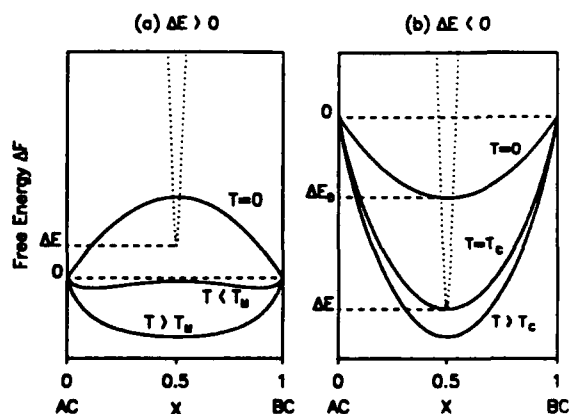


FIG. 2. Schematic plots of the mixing free energy of a disordered alloy as function of alloy composition x (dotted lines) for (a) $\Delta E > 0$ and (b) $\Delta E < 0$. The dashed line represented ΔE for an ordered alloy ABC_2 .

critical temperature T_m for phase segregation or forming miscibility gaps. However, if $\Delta E < 0$, then the ordered phase is stable at a temperature T lower than the critical temperature T_c for the order-disorder transition.

A crude estimate of the magnitude of ΔE in Fig. 2(b) required for ordering at a given temperature T can be made by assuming that the disordered phase is a random alloy. Then in the present units of energy (for an ABC_2 molecule), the magnitude of the ordering energy defined as $\Delta E_0 = |\Delta E - \Delta E_D|$ must be greater than $2kT \ln 2$, where ΔE_D is the mixing energy for the disordered phase for the case at $x = \frac{1}{2}$ and $T = 0$, also shown in Fig. 2(b). At a typical epitaxial growth temperature of 600°C (see Table I), the minimum ordering energy is estimated to be $\Delta E_0 \approx 100$ meV. Taking $\Delta E_D = \Delta E/2$, one finds a critical value of $\Delta E = -200$ meV. However, if a strict pair-potential nearest-neighbor model is used, $\Delta E_D = \frac{1}{2}\Delta E$ for both the CuAu I and chalcopyrite structures and the required ΔE value becomes -400 meV in this estimate.

Finally, consider the implication of different excess energies to ordering. Let us take Ga_2AsSb in the chalcopyrite structure as an example, for which the three excess energies are 70, -28 , and -205 meV for ΔE , ΔE_{ep} , and ΔE_{ep}^H , respectively (see Tables VI, VII, and VIII) from the TB calculation. The positive value of ΔE corresponds to the case of Fig. 2(a), which shows that the ordered bulk phase is not stable at any temperature. Although the negative value $\Delta E_{ep} = -28$ meV corresponds to the case in Fig. 2(b), the corresponding critical temperature T_c is very low. This implies a simple constraint that the lateral lattice constants of the grown material matched to the substrate lattice are not enough to produce a metastable ordered alloy at the experimental growth temperature of 600°C . However, the significant negative value of $\Delta E_{ep}^H = -205$ meV, resulting from a further constraint in which the c axis relaxation is not allowed, may raise the T_c in Fig. 2(b) into the range of the growth temperature and make the metastable ordered phase a possibility.

- ¹J. C. Mikkelsen, Jr. and J. B. Boyce, *Phys. Rev. Lett.* **49**, 1412 (1982); *Phys. Rev. B* **28**, 7130 (1983).
- ²A. Balzarotti, in *Ternary and Multinary Compounds*, Proceedings of the Seventh International Conference, edited by S. K. Deb and A. Zunger (Materials Research Society, Pittsburgh, Pennsylvania, 1987), p. 333.
- ³T. P. Pearsall, R. Carles, and J. C. Portal, *Appl. Phys. Lett.* **42**, 436 (1983).
- ⁴K. Kakimoto and T. Katoda, *Appl. Phys. Lett.* **40**, 826 (1982).
- ⁵A. Sher, M. van Schilfgaarde, A.-B. Chen, and W. Chen, *Phys. Rev. B* **36**, 4279 (1987).
- ⁶S.-W. Wei, L. G. Ferreira, and A. Zunger, *Phys. Rev. B* **41**, 8240 (1990).
- ⁷T. S. Kuan, T. F. Kuech, W. I. Wang, and E. L. Wilkie, *Phys. Rev. Lett.* **54**, 201 (1985).
- ⁸H. Nakayama and H. Fujita, in *GaAs and Related Compounds*, Proceedings of the 12th International Symposium, edited by M. Fujimoto, IOP Conf. Proc. No. 79 (Institute of Physics and Physical Society, London, 1986), p. 289.
- ⁹A. G. Norman, R. E. Mallard, I. J. Murgatroyd, G. R. Booker, A. H. Moore, and M. D. Scott, in *Microscopy of Semiconducting Materials*, edited by A. G. Gullis and P. A. Augustus, IOP Conf. Proc. No. 87 (Institute of Physics and Physical Society, London, 1987), p. 77.
- ¹⁰Y. Matsui, H. Hayashi, and K. Yoshida, *Appl. Phys. Lett.* **48**, 1060 (1986).
- ¹¹T. S. Kuan, W. I. Wang, and E. L. Wilkie, *Appl. Phys. Lett.* **51**, 51 (1987).
- ¹²M. A. Shahid, S. Mahajan, D. E. Laughlin, and H. M. Cox, *Phys. Rev. Lett.* **58**, 2567 (1987). Note that these authors used the vapor-levitation-epitaxy (VLE) method.
- ¹³S. Yasuami, C. Nozaki, and Y. Ohba, *Appl. Phys. Lett.* **52**, 2031 (1988).
- ¹⁴A. Gomoyo, T. Suzuki, K. Kobayashi, S. Kawata, I. Hino, and T. Yuasa, *Appl. Phys. Lett.* **50**, 673 (1987); A. Gomoyo, T. Suzuki, and S. Iijima, *Phys. Rev. Lett.* **60**, 2645 (1988).
- ¹⁵P. Bellon, J. P. Chevalier, G. P. Martin, E. Dupont-Nivet, C. Thiebaud, and J. P. André, *Appl. Phys. Lett.* **52**, 567 (1988).
- ¹⁶O. Ueda, M. Takikawa, J. Komeno, and I. Umebu, *Jpn. J. Appl. Phys.* **26**, L1824 (1987).
- ¹⁷M. Kondow, H. Kakibayashi, and S. Minagawa, *J. Cryst. Growth* **88**, 291 (1988).
- ¹⁸H. R. Jen, M. J. Cherng, and G. B. Stringfellow, *Appl. Phys. Lett.* **48**, 1603 (1986).
- ¹⁹I. J. Murgatroyd, A. G. Norman, G. R. Booker, and T. M. Kerr, in *Proceedings of the Eleventh International Congress on Electron Microscopy, Kyoto, Japan, 1986*, edited by T. Imura, S. Maruse, and T. Suzuki (Japan Society of Electron Microscopy, Tokyo, 1986), p. 1497.
- ²⁰S.-H. Wei and A. Zunger, *Phys. Rev. B* **39**, 6279 (1989).
- ²¹G. B. Stringfellow, *J. Cryst. Growth* **27**, 21 (1974); *J. Appl. Phys.* **54**, 404 (1983).
- ²²M. B. Panish and M. Ilegems, *Prog. Solid State Chem.* **7**, 39 (1972).
- ²³V. T. Bublik and V. N. Leikin, *Phys. Status Solidi A* **46**, 365 (1978).
- ²⁴R. Patrick, A.-B. Chen, and A. Sher, *Phys. Rev. B* **36**, 6585 (1987).
- ²⁵The magnitude of the excess energy $|\Delta E|$ required to achieve an order-disorder transition at a given transition temperature T_c varies from model to model [see D. de Fontaine, in *Solid State Physics: Advances in Research and Applications*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1979), Vol. 34, p. 73, particularly Table V], with the value $|\Delta E|_R$ for the regular-solution model being the smallest. The value of $|\Delta E|_R$ for the CuAu structure in the present scale is $3kT_c = 225$ MeV per ABC_2 molecule at $T_c = 900$ K. The actual $|\Delta E|$ should be larger than $|\Delta E|_R$.
- ²⁶A.-B. Chen and A. Sher, *Phys. Rev. B* **32**, 3695 (1985).
- ²⁷J. E. Bernard, L. G. Ferreira, S.-H. Wei, and A. Zunger, *Phys. Rev. B* **38**, 6338 (1988).
- ²⁸P. Boguslawski and A. Baldereschi, *Phys. Rev. B* **39**, 8055 (1989).
- ²⁹T. Ohno, *Phys. Rev. B* **38**, 13 191 (1988).
- ³⁰A. Qreish, N. Motta, and A. Balzarotti, *Phys. Rev. B* **39**, 5987 (1989).
- ³¹D. M. Bylander and L. Kleinman, *Phys. Rev. B* **34**, 5280 (1986); **36**, 3229 (1987).
- ³²G. P. Srivastava, J. L. Martins, and A. Zunger, *Phys. Rev. B* **38**, 12 694 (1988).

- ³³M. Methfessel, C. O. Rodriguez, and O. K. Andersen, Phys. Rev. B **40**, 2009 (1989).
- ³⁴E. Wimmer, H. Krakauer, M. Weinert, and A. J. Freeman, Phys. Rev. B **24**, 864 (1981).
- ³⁵G. B. Bachelet, D. R. Hamman, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
- ³⁶P. N. Keating, Phys. Rev. **145**, 637 (1966).
- ³⁷R. M. Martin, Phys. Rev. B **1**, 4005 (1970).
- ³⁸D. J. Chadi, Phys. Rev. B **19**, 2074 (1979); **29**, 785 (1984).
- ³⁹W. A. Harrison, *Electronic Structure and Properties of Solids* (Freeman, San Francisco, 1980).
- ⁴⁰W. A. Harrison, Phys. Rev. B **27**, 3592 (1983).
- ⁴¹J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
- ⁴²L. Kleinman, Phys. Rev. **128**, 2614 (1962).
- ⁴³A.-B. Chen and A. Sher, Phys. Rev. B **31**, 6490 (1985), Table I.
- ⁴⁴R. Zallen, in *Handbook on Semiconductors*, edited by T. S. Moss (North-Holland, Amsterdam, 1982), Vol. 1, Chap. 1.
- ⁴⁵J. C. Phillips, *Bonds and Bands in Semiconductors* (Academic, New York, 1973).
- ⁴⁶*Numerical Data and Functional Relationship in Science and Technology*, edited by O. Madelung, Landolt-Börnstein, New Series, Group 3, Vol. 17 (Springer-Verlag, Berlin, 1982); *ibid.*, Vol. 22 (1988).
- ⁴⁷All bond lengths are deduced from the lattice constants quoted by Ref. 44. The values of E_{bond} are taken from Table 7-3 in Ref. 39, except for AlSb, ZnTe, CdTe, and HgTe, which are deduced from Table 8.2 in Ref. 45. The elastic constants and the phonon frequencies are taken from values compiled in Ref. 46.
- ⁴⁸Note that the parameters u_0 , u_1 , and u_2 and the calculated elastic constants based on Chadi's TB parameters in this table are not the same as Chadi's published (1979) values, which give bulk moduli about 20% smaller than the experimental values by considerably better phonon frequencies.
- ⁴⁹O. H. Nielsen and R. M. Martin, Phys. Rev. Lett. **50**, 697 (1983); Phys. Rev. B **32**, 3792 (1985); **32**, 3780 (1985).
- ⁵⁰C. S. G. Cousins, L. Staun, J. Staun, B. Selsmark, and B. J. Sheldon, J. Phys. C **20**, 29 (1987).
- ⁵¹D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973).
- ⁵²S. Froyen, Phys. Rev. B **39**, 3168 (1989).
- ⁵³J. E. Bernard, R. G. Dandrea, L. G. Ferreira, S. Froyen, S.-H. Wei, and A. Zunger, Appl. Phys. Lett. **56**, 731 (1990).
- ⁵⁴J. S. Nelson and I. P. Batra, Phys. Rev. B **39**, 3250 (1989).
- ⁵⁵B. I. Min, S. Massidda, and A. J. Freeman, Phys. Rev. B **38**, 1970 (1988).
- ⁵⁶S. Ciraci and I. P. Batra, Phys. Rev. Lett. **58**, 14 (1987); Phys. Rev. B **36**, 1225 (1987).
- ⁵⁷C. P. Flynn, Phys. Rev. Lett. **57**, 599 (1986).
- ⁵⁸J. B. Boyce and J. C. Mikkelsen, in *Ternary and Multinary Compounds*, Proceedings of the Seventh International Conference, edited by S. K. Deb and A. Zunger (Materials Research Society, Pittsburgh, Pennsylvania, 1987), p. 359.
- ⁵⁹W.-F. Pong, R. A. Mayanovic, B. A. Bunker, J. K. Furdyna, and U. Debska, Phys. Rev. B **41**, 8440 (1990).
- ⁶⁰L. G. Ferreira, S.-H. Wei, and A. Zunger, Phys. Rev. B **40**, 3197 (1989).
- ⁶¹Based of the VFF analysis, Martin in Ref. 37 shows that B scales as $1/d^3$, while in an empirical formula by M. L. Cohen [Phys. Rev. B **32**, 7988 (1985)] B scales as $1/d^{3.5}$.
- ⁶²A.-B. Chen, A. Sher, and M. A. Berding, Phys. Rev. B **37**, 6285 (1988).
- ⁶³G. P. Srivastava, J. L. Martins, and A. Zunger, Phys. Rev. B **31**, 2561 (1985).
- ⁶⁴J. S. Cohen and A. G. Schlijper, Phys. Rev. B **36**, 1526 (1987).
- ⁶⁵W. I. Wang, Appl. Phys. Lett. **58**, 3244 (1985).
- ⁶⁶M. van Schilfgaarde, A.-B. Chen, and A. Sher, Phys. Rev. Lett. **57**, 1149 (1986).
- ⁶⁷J. S. Cohen and A. G. Schlijper, Phys. Rev. B **38**, 12694 (1988).
- ⁶⁸A.-B. Chen, M. van Schilfgaarde, and A. Sher, Bull. Am. Phys. Soc. **31**, 664 (1986).
- ⁶⁹D. M. Wood, S.-H. Wei, and A. Zunger, Phys. Rev. B **37**, 1342 (1988).
- ⁷⁰S.-H. Wei, Phys. Rev. Lett. **59**, 2613 (1987).
- ⁷¹The negative ΔE values obtained from LDF in Refs. 63 and 64 were revised in Refs. 67 and 54, respectively. The negative Madelung energy in the ordered alloy proposed in Ref. 65 only considered the contributions from the alloying species. It was pointed out in Refs. 68 and 69 that if all species are considered, the dipolar energy for both ordered and disordered alloys are positive. Finally, the suggestion in Ref. 68 that the total Coulomb energy for a disordered alloy might exceed the ordered system was also revised in Ref. 70.
- ⁷²S. Krishnamurthy, M. A. Berding, A. Sher, and A.-B. Chen, Phys. Rev. Lett. **64**, 2531 (1990).
- ⁷³J. E. Bernard, R. G. Dandrea, L. G. Ferreira, S. Froyen, S.-H. Wei, and A. Zunger, Bull. Am. Phys. Soc. **35**, 667 (1990).
- ⁷⁴R. G. Dandrea, J. E. Bernard, S.-H. Wei, and A. Zunger, Phys. Rev. Lett. **64**, 36 (1990).

Appendix D

**ELASTIC CONSTANTS AND RELATED PROPERTIES OF
SEMICONDUCTOR COMPOUNDS AND THEIR ALLOYS**

**ELASTIC CONSTANTS AND RELATED PROPERTIES
OF SEMICONDUCTOR COMPOUNDS AND THEIR ALLOYS**

A.-B. Chen

Physics Department, Auburn University, Auburn, AL 36849

A. Sher

Physical Electronics Laboratory, SRI International, Menlo Park, CA 94025

W.T. Yost

NASA Langley Research Center, Hampton, VA 23665

CONTENTS

I.	INTRODUCTION.....	1
1.	Goals	1
2.	Definition and Calculation of Elastic Constants	1
3.	Elastic Constants and Sonic Wave Propagation	5
II.	MEASUREMENT METHODS	12
4.	Velocity Measurements.....	12
5.	Ultrasonic Measurement Techniques.....	13
6.	Corrections to Pulse Measurements: An Example	16
7.	Optical Techniques.....	20
III.	THEORETICAL AND EXPERIMENTAL RESULTS	25
8.	AB INITIO Theory	25
9.	Valence Force Field Model.....	29
10.	Tight Binding Theory.....	33
11.	Semiconductor Alloys	46
IV.	DISLOCATIONS AND HARDNESS	57
12.	Slip Systems.....	58
13.	Peierls Energy	59
14.	Temperature Dependence.....	62
V.	CONCLUDING REMARKS	64
	REFERENCES	66

ACKNOWLEDGMENTS

This work was supported in part by AFOSR Contract F49620-88-K-0009 and ONR Contract N00014-88-C0096. We are especially indebted to M.A. Berding for permitting us to include in this paper unpublished ideas from a segment of a report she wrote. We also wish to acknowledge contributions from J. Heyman.

I. INTRODUCTION

1. GOALS

The goals of this paper are to review the current state of knowledge of the elastic constants of elemental, compound, and pseudobinary alloy semiconductors. To accomplish this objective, we will review:

- Experimental methods currently used to measure elastic constants,
- Experimental results
- Binding and elastic constant theory
- Related mechanical properties

Heavy emphasis is placed on comparisons between theory and experiment, and the accuracy of approximations currently in vogue. The theories discussed range from first-principles methods, requiring heavy computations, to parameterized physical models. The intent is to identify a logical path between these extremes, and thereby provide insight into the connection between atom potentials and semiconductor mechanical properties. In the course of this presentation, there are a number of instances where the need for improvements in the theory, or for additional experimental results, is encountered. We have tried to highlight these situations and suggest possible remedies.

2. DEFINITION AND CALCULATION OF ELASTIC CONSTANTS

The theory of elasticity of solids has been well formulated in many treatises (Love, 1944; Landau and Lifshitz, 1959), so we need not review the different formalisms and conventions.

However, as we do wish to present a coherent account of the essence of the theory, it is necessary to define terms and describe general calculations to be used later.

In linear elasticity theory, deformation is assumed to be infinitesimal. The relative displacement vector \vec{x} between two points in a deformed solid is related to the corresponding vector \vec{x} in the undeformed solid by the following equation, in component form, through a 9-component strain tensor ϵ :

$$x'_\alpha = x_\alpha + \sum_\beta \epsilon_{\alpha\beta} x_\beta \quad . \quad (1)$$

The change in the internal energy associated with ϵ is also small and will be denoted UV , where V is the equilibrium volume of a solid. Under the condition that the entropy and the electrostatic displacement field are constant, U is only a function of ϵ , and is quadratic in ϵ :

$$U = \frac{1}{2} \sum_{\alpha\beta\mu\nu} \epsilon_{\alpha\beta} C_{\alpha\beta\mu\nu} \epsilon_{\mu\nu} \quad . \quad (2)$$

where $C_{\alpha\beta\mu\nu}$ are the elastic stiffness coefficients, which are characteristic properties of the solid. From this definition, $C_{\mu\nu\alpha\beta} = C_{\alpha\beta\mu\nu}$ is required for U to be an analytical function of ϵ . Moreover $C_{\alpha\beta\mu\nu} = C_{\beta\alpha\mu\nu} = C_{\alpha\beta\nu\mu}$ is also required to ensure that U is zero under any infinitesimal rigid rotation. In light of these properties, the energy can be expressed in terms of a symmetrical strain tensor η as

$$U = \frac{1}{2} \sum_{\alpha\beta\mu\nu} \eta_{\alpha\beta} c_{\alpha\beta\mu\nu} \eta_{\mu\nu} \quad . \quad (3)$$

where $\eta_{\alpha\beta}$ is defined as

$$\eta_{\alpha\beta} = \frac{1}{2} (\epsilon_{\alpha\beta} + \epsilon_{\beta\alpha}) \quad . \quad (4)$$

The strain tensor η is a thermodynamic parameter with stress tensor σ as its conjugate variable (Bragger, 1964). The components of σ are given by

$$\sigma_{\alpha\beta} = \frac{\partial U}{\partial \eta_{\alpha\beta}} = \sum_{\mu\nu} C_{\alpha\beta\mu\nu} \eta_{\mu\nu} \quad (5)$$

It is clear that σ is also a symmetrical tensor.

The most frequently used notation is the engineering convention, in which the strain tensor e is related to ϵ of Eq. (4) by $e_{\alpha\alpha} = \eta_{\alpha\alpha}$ for the diagonal components, but $e_{\alpha\beta} = 2\eta_{\alpha\beta}$ for $\alpha \neq \beta$. Furthermore, since e has at most six independent components, it is treated as a 6-component vector, with the vector components 1 to 6 corresponding respectively to the tensor components xx , yy , zz , yz , xz , and xy . In terms of e the strain energy density is written as

$$U = \frac{1}{2} \sum_{ij} C_{ij} e_i e_j \quad (6)$$

where C_{ij} can be identified as $c_{\beta\alpha\mu\nu}$ with $i = \alpha\beta$ and $j = \mu\nu$. Because C_{ij} is symmetrical, it has at most 21 independent components for any crystal. Crystal symmetries reduce this number further (Love, 1944, Ashcroft and Mermin, 1976). For a cubic lattice, to which class the zincblende and diamond semiconductors belong, there are only three independent components, namely C_{11} , C_{12} , and C_{44} .

The three independent elastic constants of the diamond and zincblende (zb) semiconductors can be calculated by considering the following three strains.

(i) Under a uniform expansion, which changes a displacement \vec{x} into $\vec{x}' = (1 + e) \vec{x}$, then $e_1 = e_2 = e_3 = e$ and the other strain components are zero; the elastic energy density is then given by

$$U = 3(C_{11} + 2C_{12})e^2/2 \quad (7)$$

U can also be expressed in terms of the adiabatic bulk modulus B defined by $\delta P = -B(\delta V/V)$, where V is the crystal volume, δV is its change and δP is the corresponding pressure change. The result is $U = B(\delta V/V)^2/2 = 9Be^2/2$, because the dilatation is $\delta V/V = 3e$ in the present case. This establishes the relationship

$$B = (C_{11} + 2C_{12})/3 \quad (8)$$

(ii) The next case to consider is a tetragonal shear strain e which changes a displacement according to

$$(x, y, z) \rightarrow (x + ex, y - ey, z) . \quad (9)$$

The only nonzero strain components are $e_1 = -e_2 = e$. Then U becomes

$$U = (C_{11} - C_{12}) e^2 . \quad (10)$$

(iii) To calculate C_{44} we consider a shear strain e which changes a displacement according to

$$(x, y, z) \rightarrow (x + ey/2, y + ex/2, z) . \quad (11)$$

This strain contains e_6 as the only nonzero component, Because C_{44} equals C_{66} , the energy density is simply

$$U = C_{44} e^2/2 . \quad (12)$$

Although the macroscopic crystal distortion of the Bravais lattice caused by this strain is described by Eq. (11), microscopically there is a relative displacement $\vec{u} = (0,0,u)$, the so-called Kleimann (1962) internal displacement, between two successive atomic planes perpendicular to the z axis. In other words, the relative displacements between the atoms on the same fcc sublattice are governed by Eq. (11), but there is an additional induced relative displacement \vec{u} between the two sublattices. The directions of the displacements of atoms in a tetrahedral cell are shown in Figure 1. In calculations one can use an arbitrary infinitesimal pair of e and u to obtain the coefficients in the following quadratic expansion of the strain energy density:

$$U = \Phi u^2/2 + D eu + C_{44}^{(o)} e^2/2 , \quad (13)$$

where the force function Φ is related to the transverse optical phonon frequency ω at Γ (the center of Brillouin zone) by $\Phi = \mu \omega^2$ with μ being the reduced mass. In Eq. (12), D is a constant, and $C_{44}^{(o)}$ would be the shear stiffness coefficient if the internal displacement were not allowed.

Kleimann (1962) defined an internal displacement parameter which is related to the equilibrium value of u by $u = \zeta ae/4$ for a fixed e , where a is the lattice constant. Taking the first derivative of U with respect to u in Eq.(13) and setting it equal to zero, one finds that the ζ value is given by

$$\zeta = -4D/(a\Phi) \quad . \quad (14)$$

Finally the sought-after C_{44} is given by

$$C_{44} = C_{44}^{(0)} - \zeta^2 a^2 \Phi / 16 \quad . \quad (15)$$

Thus the internal displacement is an essential part of C_{44} . These procedures will be used in the theoretical calculations to be presented in subsequent sections.

3. ELASTIC CONSTANTS AND SONIC WAVE PROPAGATION

The purpose of this section is to give the reader sufficient information to measure ultrasonic wave velocities and to deduce from them the elastic coefficients of materials of interest to the electronics industry. We start from the definition of strain and develop equations that relate the elastic coefficients with wave propagation velocities in various crystallographic directions. The cases taken as examples, cubic and isotropic, were chosen because of their prevalent role in electronic materials.

A number of techniques used to measure wave propagation velocities in materials are discussed. The researcher's decision on technique depends upon a number of factors, such as accuracy needed, size and crystallographic orientation of available samples, equipment, etc. Sample preparation is significant since a carelessly prepared sample yields useless velocity data. Because the equipment needs vary widely among the techniques that can be used, references containing information on each technique are given. Key papers chosen for citation were picked for their readability and their direct application to the measurement technique. Primary emphasis has been given to techniques that are generally accepted, and that give results that are understood. Throughout this section a general theory of elasticity due to Murnaghan (1951) is used.

a. Elastic to Wave Relationship

Consider, as before, a point in a lossless elastic medium at rest, whose displacement vector, $\vec{x} = (x_\alpha)$, $\alpha = 1, 2, 3$ are the cartesian coordinates of a point in the medium. In the presence of a stress wave disturbance, the point moves to a new location, $\vec{x}' = (x'_\alpha)$ at time t . The Lagrangian strains, η , defined consistently with Eqs. (1) and (4) but with higher order terms retained (Landau and Lifshitz, 1986; Murnaghan, 1951) are as follows:

$$\begin{aligned}\eta_{\mu\nu} &= \frac{1}{2} \left(\frac{\partial x'_\alpha}{\partial x_\mu} \frac{\partial x'_\alpha}{\partial x_\nu} - \delta_{\mu\nu} \right) \\ &= \frac{1}{2} \left(\frac{\partial u_\mu}{\partial x_\nu} + \frac{\partial u_\nu}{\partial x_\mu} + \frac{\partial u_\alpha}{\partial x_\mu} \frac{\partial u_\alpha}{\partial x_\nu} \right)\end{aligned}\quad (16)$$

where $\delta_{\mu\nu}$ is the Kronecker delta. Einstein summation convention over repeated indices is used in this section.

By considering the internal energy per unit mass of the medium $E(r_\alpha, \eta_{\alpha\beta}, S)$, where S is the entropy, we obtain a relationship between wave propagation and the internal energy of the medium. If we restrict this discussion to a medium that is initially unstressed, then the internal energy per unit volume of the medium is $\rho_0 E(r_\alpha, \eta_{\alpha\beta}, S)$, where S_0 is the mass density of the medium in the unstrained state. By expanding this in a power series of the Lagrangian strains, we obtain

$$\rho_0 E(r_\alpha, \eta_{\alpha\beta}, S) = \rho_0 E(r_\alpha, 0, S) + \frac{1}{2!} C_{\mu\nu\alpha\beta} \eta_{\mu\nu} \eta_{\alpha\beta} + \frac{1}{3!} C_{\mu\nu\alpha\beta\gamma\delta} \eta_{\mu\nu} \eta_{\alpha\beta} \eta_{\gamma\delta} + \dots \quad (17)$$

The coefficients $C_{\mu\nu\alpha\beta}$ and $C_{\mu\nu\alpha\beta\gamma\delta}$ are the second-order and third-order elastic coefficients (adiabatic) as defined by Brugger (1964):

$$C_{\mu\nu\alpha\beta} = \rho_0 \left(\frac{\partial^2 E}{\partial \eta_{\mu\nu} \partial \eta_{\alpha\beta}} \right), \quad (18)$$

$$C_{\mu\nu\alpha\beta\gamma\delta} = \rho_0 \left(\frac{\partial^3 E}{\partial \eta_{\mu\nu} \partial \eta_{\alpha\beta} \partial \eta_{\gamma\delta}} \right) . \quad (19)$$

The wave equation can be derived (Goldstein, 1965, pp. 347) from Lagrange's equation

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_\mu} \right) + \frac{d}{dx_\alpha} \left(\frac{\partial L}{\partial \left(\frac{\partial u_\mu}{\partial x_\alpha} \right)} \right) - \frac{\partial L}{\partial u_\mu} = 0 , \quad (20)$$

where $\dot{x}_\mu = \partial x_\mu / \partial t$ and L is given by

$$L = \frac{1}{2} \rho_0 \dot{u}_\mu \dot{u}_\mu - \rho_0 E(x_\alpha, \eta_{\alpha\beta}, S) . \quad (21)$$

Defining the particle displacement vector, u , as before

$$u_\mu(\vec{r}, t) = \dot{x}_\mu - x_\mu , \quad (22)$$

and differentiating it gives

$$\frac{\partial u_\mu}{\partial x_\nu} = \frac{\partial \dot{x}_\mu}{\partial x_\nu} - \delta_{\mu\nu} . \quad (23)$$

Combining Eqs. (22), (23), and (16) gives the strain tensor in terms of the particle displacements,

$$\eta_{\mu\nu} = \frac{1}{2} \left(\frac{\partial u_\mu}{\partial x_\nu} + \frac{\partial u_\nu}{\partial x_\mu} + \frac{\partial u_\beta}{\partial x_\mu} \frac{\partial u_\beta}{\partial x_\nu} \right) . \quad (24)$$

Using Eqs. (21) to (24) in Eq. (20), and retaining first order terms in $\partial u_\mu / \partial x_\nu$, gives

$$\rho_0 \ddot{u}_\mu = \frac{d}{dx_\nu} \left(C_{\mu\nu\alpha\beta} \frac{\partial u_\alpha}{\partial x_\beta} \right) . \quad (25)$$

We assume a plane wave solution to Eq. (25) of the form

$$u_\mu = \hat{e}_\mu \cdot \vec{u} \cos(k_\nu x_\nu - \omega t) \quad (26)$$

where

\hat{e}_μ = unit displacement vector in the μ th direction

$\vec{u} = (u_\mu)$ = particle displacement vector

$\vec{k} = (k_\nu)$ = wave propagation vector

ω = angular frequency.

Substituting Eq. (26) into Eq. (25), we obtain the eigenvalue-eigenvector equation

$$|C_{\mu\nu\alpha\beta} \kappa_\alpha \kappa_\beta - \rho_0 v^2 \delta_{\mu\alpha}| = 0 \quad (27)$$

where

(κ_μ) = the set of direction cosines of $\vec{\kappa}$

$v = \omega/|\vec{k}|$ is the phase speed

$\vec{\kappa} = \vec{k}/|\vec{k}|$, indicates the direction of wave propagation in the medium.

For any given wave propagation direction, there are three eigenvalue solutions to Eq. (27); these correspond to one quasilongitudinal and two quasitransverse polarization modes. Relationships have also been derived for wave propagation in the presence of residual stresses and for retention of the higher-order elastic coefficients (Cantrell, 1982; Breazeale and Ford, 1965).

Because both the stress and strain tensors are symmetric, one can reduce the number of independent second-order elastic coefficients to 21 (Landau and Lifshitz, 1986, p. 32). It is also convenient to use the Voight (1928) contraction of the indices as in Eq. (6) which is used in the remainder of this section. The number of independent elastic coefficients varies with crystal structure¹, and is listed below.

Crystal Class	Number of Independent Elastic Coefficients
Triclinic	21
Monoclinic	13
Orthorhombic	9
Tetragonal (C_4, S_4, C_{4h})	7
Tetragonal ($C_{4v}, D_{2d}, D_4, D_{4h}$)	6
Rhombohedral (C_3, S_6)	7
Rhombohedral (C_{3v}, D_3, D_{3d})	6
Hexagonal	5
Cubic	3
(Isotropic)	2

Green (1973) and others have solved Eq. (27) for isotropic materials and for cubic crystals in the [100], [110], and [111] directions. Using the symmetry properties of the crystals, one obtains the results presented in subsection b for cubic crystals, and in subsection c for isotropic solids.

b. Cubic Crystals

For plane waves propagating along the [100] direction:

$$(\kappa_1 = 1, \kappa_2 = 0, \kappa_3 = 0) ;$$

as longitudinal (compressional) waves, pure mode (particle displacements: u_1 in direction of propagation, u_2 and $u_3 = 0$)

$$v_1 = \sqrt{\frac{C_{11}}{\rho_0}} ; \quad (28)$$

or as transverse (shear) waves, pure mode (particle displacements: $u_1 = 0$, u_2 and u_3 are perpendicular to the direction of propagation)

$$v_2 = v_3 = \sqrt{\frac{C_{44}}{\rho_0}} . \quad (29)$$

For plane waves propagating along the [110] direction:

$$(\kappa_1 = 1/\sqrt{2}, \kappa_2 = 1/\sqrt{2}, \kappa_3 = 0) ;$$

as longitudinal (compressional) waves, pure mode (particle displacements: $u_1 = u_2, u_3 = 0$)

$$v_1 = \sqrt{\frac{(C_{11} + C_{12} + 2C_{44})}{2\rho_0}} ; \quad (30)$$

as transverse (shear) waves, pure mode ($u_1 = -u_2, u_3 = 0$)

$$v_2 = \sqrt{\frac{(C_{11} - C_{12})}{2\rho_0}} ; \quad (31)$$

or as transverse (shear) waves, pure mode ($u_1 = u_2 = 0, u_3 \neq 0$)

$$v_3 = \sqrt{\frac{(C_{44})}{2\rho_0}} . \quad (32)$$

For plane waves propagating along the [111] direction:

$$(\kappa_1 = 1/\sqrt{3}, \kappa_2 = 1/\sqrt{3}, \kappa_3 = 1/\sqrt{3}) ; \quad (33)$$

as longitudinal (compressional) waves, pure mode ($u_1 = u_2 = u_3$)

$$v_1 = \sqrt{\frac{(C_{11} + 2C_{12} + 4C_{44})}{3\rho_0}} ; \quad (34)$$

or as transverse (shear) waves, pure mode ($u_1\kappa_1 + u_2\kappa_2 + u_3\kappa_3 = 0$ or particle displacement perpendicular to wave propagation)

$$v_2 = v_3 = \sqrt{\frac{(C_{11} - C_{12} + C_{44})}{3\rho_0}} . \quad (35)$$

c. Isotropic Solids

For the case of isotropy, all directions are equivalent, so we are left with two independent constants (Lame' constants). The elastic coefficients can be expressed in terms of the Lame' constants (Green, 1973) as follows.

$$\begin{aligned}
C_{11} &= C_{22} = C_{33} = \lambda + 2\mu \\
C_{12} &= C_{13} = C_{23} = \lambda \\
C_{21} &= C_{31} = C_{32} = \lambda \\
C_{44} &= C_{55} = C_{66} = \mu
\end{aligned}
\tag{36}$$

For plane waves propagating in any direction (e.g., the x-direction):

$$(\kappa_1 = 1, \kappa_2 = 0, \kappa_3 = 0) \text{ .}$$

With longitudinal (compressional) wave, pure mode (particle displacements: u_1 in direction of propagation, u_2 and $u_3 = 0$)

$$v_1 = \sqrt{\frac{\lambda + 2\mu}{\rho_0}} \text{ .} \tag{37}$$

With transverse (shear) waves, pure mode (particle displacements: $u_1 = 0$, and u_2 and u_3 perpendicular to the direction of propagation)

$$v_2 = v_3 = \sqrt{\frac{\mu}{\rho_0}} \text{ .} \tag{38}$$

II. MEASUREMENT METHODS

4. VELOCITY MEASUREMENTS

To measure the second-order elastic coefficients (SOEC), one can determine the sound velocity and the density of the sample, and calculate the combination of SOEC that the direction of propagation requires. For cubic systems, sound velocities measured in the pure mode directions equivalent to [100], [110], and [111] make it possible to determine the three independent elastic constants, C_{11} , C_{12} , and C_{44} . Using care in sample preparation, and appropriate corrections for bond thickness and diffraction, the sound velocities can be determined to parts in 10^5 . The accuracy of a velocity determination is usually limited by the accuracy of the path length measurement in pulsed and continuous wave techniques. With optical techniques, deflection angles, wavelength measurements, and frequencies are determinant factors.

In general, there is little problem in obtaining a sound velocity measurement on the order of 1% uncertainty. But to improve on this, one must exercise additional care in the preparation of surfaces, the control of temperature, the determination of acoustic path length, and in the determination of travel time of the acoustic wave². For single transducer configurations, where a transducer is used to send and receive the acoustic wave, the base equation, for determining the velocity using pulsed techniques, or pulse echo, without corrections, is as follows:

$$v = \frac{2L}{T} \quad , \quad (39)$$

where v is the wave velocity, L is the sample length, and T is the round-trip time. Corrections³ are made in the evaluation of T , and vary with the technique used in the measurement. Corrections for bond thickness (McSkimin, 1961), and for diffraction effects (Papadakis, 1961) can be measured and/or calculated. Accuracies of some of the more frequently used pulsed techniques

are given in Table 1. A comparison of the accuracies of various techniques are discussed by Papadakis (1972, 1976). Using continuous wave techniques, the propagating plane wave model, the basic equation without corrections is

$$v = 2L \Delta f \quad (40)$$

where Δf is the frequency difference between two adjacent mechanical resonance modes. Typically, velocity measurement accuracy is good to about 1% to 10%. A correction factor (Chern, 1981) can be applied to the right side of Eq. (39) that corrects for the effects of the transducer and bond on the mechanical resonances of the sample. When applied, the inaccuracy can be as low as 5 parts per 100,000, neglecting inaccuracies in path length measurements.

If the sample is transparent, optical techniques offer a convenient method to measure sound velocity. For some of these techniques, the accuracy is on order of parts per thousand. Still, some offer accuracies of parts per ten thousand and can be used with small samples, and very high frequency ultrasonic waves (Breazeale et al., 1981). In general the calculation for velocity involves the determination of acoustic wavelength in the medium, by using the equation

$$v = f\lambda \quad (41)$$

where f is the frequency of the sound wave, determined by the drive frequency of the transducer, and λ is the wavelength. Brillouin scattering also makes it possible to measure the sound velocity (Beyer and Letcher, 1969). In addition to measurement of an angle, the accuracy also depends on determination of the wavelength of light in the medium; this requires a determination of the index of refraction of the medium.

5. ULTRASONIC MEASUREMENT TECHNIQUES

A survey of techniques and details for measuring ultrasonic velocities are available from various sources (Breazeale et al., 1981, Papadakis, 1976, Truell et al., 1969). Consideration in

this work will be given to three classes of techniques for absolute velocity measurements. A compilation of some of the features of each technique are listed in Table 2.

a. Sample Preparation

The tolerance selected for preparing the sample depends largely upon the accuracy needed in the determination. If a 1% measurement is in order, surface preparation and parallelism are less critical than if an accuracy of 0.01% is desired. Also important is the consideration of correction factors needed to compensate for bond thickness and diffraction effects. At present diffraction corrections exist only for compressional waves. Therefore, under equivalent experimental conditions, the most accurate determination of combinations of elastic coefficients would be those that are calculated from compressional wave velocities.

It is assumed that a surface on the sample has been ground to optical tolerances, and that the crystallographic direction of the axis of this surface has been determined. Typically, the crystallographic directions are measured to within minutes of arc (McSkimin and Andreatch, 1964). The remaining critical issues are the parallelism of the reflecting surface to the reference surface, and the flatness of these surfaces. Parallelism of the surfaces can be measured, for example, with an autocollimator or a He-Ne gas laser and some mirrors. By placing the sample's reference surface on a stationary flat, dust-free surface, rotating the sample, and measuring the diameter of the circle traced by the beam reflected from the sample's top surface, and measuring the total path length of the beam from the sample to the image screen, one can determine the parallelism of the surfaces. The surfaces should be parallel to an angle better than $0.01 \lambda_{\text{acoustic}} / \text{transducer diameter}$. For a typical 1/2 inch diameter, 10 MHz transducer and a typical solid, the surfaces must be parallel to better than 4×10^{-4} radians (1 degree, 23 minutes). Higher frequencies, multiple reflections, and continuous wave techniques require proportionally smaller tolerances.

Typically, one also tries for a flatness of better than 1/100 of an acoustic wavelength for accurate determination of transit times. If the resonant frequency of the transducer is 10 MHz,

then the wavelength in a solid is in the neighborhood of 5×10^{-4} meters, which puts the flatness requirement near 5 micrometers. Using an optical flat and Newton's rings analysis, one can determine the flatness of the sample. For higher frequency transducers and continuous wave techniques, the number of fringes are appropriately decreased. At 100 MHz, pulsed mode, for example, the tolerance is down to less than several fringes. By using flat lapping surfaces, this is easily achieved. When multiple reflections are involved in the measurement, one must consider that for each reflection, changes occur in the wavefront direction; the phase change across the surface of the transducer at each reflection is due to lack of parallelism between the faces. For a case where a 20 MHz transducer was used, and multiple reflections employed for the measurement, Papadakis (1967) quotes a sample surface parallelism and flatness of better than 10^{-4} inches per inch.

b. Piezoelectric Transducers

Selection of the piezoelectric transducer depends largely upon the wave mode, the electronic equipment to be used for the measurement, and the personal preference of the researcher. In general, piezoelectric crystals⁴ such as quartz (high electrical impedance) and lithium niobate, or poled ceramics such as PZT (lead zirconate titanate), are chosen. Their physical and electrical properties are covered elsewhere (O'Donnell et al., 1981). Other piezoelectric transducers, including polymeric materials⁵, are also available. The transducer diameter should be smaller than the sample to assure that propagation modes are not affected by the location of lateral boundaries (Tu et al., 1955). This is especially important where measurements are taken on small samples which may have their lateral sides close to each other. One also must consider the problem as it relates to wave propagation, since wave modes other than those considered here can be excited in materials of small dimensions. From studies on cylindrical specimens, the minimum sample diameter can be no less than approximately 2.5λ , where λ is the wavelength of the ultrasound in the medium. The transducer diameter is chosen after the frequency is selected. Generally, the transducer diameter should be no larger than half the diameter of the sample, to

prevent any interference in the measurement from reflections off the lateral boundaries caused by diffraction effects.

Selection of a bonding material depends upon the type of wave (compressional or shear), and the temperature range through which the measurements are to be taken. At room temperature, a good choice for both types of waves is Salol or phenyl salicylate benzoate (Papadakis, 1964), while Dow-Corning DC-200 silicone is good for longitudinal waves. At other temperatures (McSkimin, 1957), other bonding materials, such as Nonaq stopcock grease, are used for compressional and shear waves. Various resins can be used with shear wave transducers; commercially prepared bonding materials are also available⁶.

Application of the bond requires care to keep surfaces clean and free of dust. The bond should be as thin as practical, taking care not to break the transducer. Often, applying some heat helps with viscous bonding materials. When using Salol or phenyl benzoate for bonds, certain procedures can be followed to assure uniform bonds (Papadakis, 1964).

c. Noncontacting Transducers

There are three types of noncontacting transducers: capacitive transducers, (compressional wave excitation only) (Cantrell and Breazeale, 1977), electromagnetic transducers (EMAT) (Vasile and Thompson, 1977; Johnson and Mase, 1984), and optically stimulated acoustic transducers (Prosser and Green, 1985). With the exception of the electrostatic transducer, (compressional with circular piston geometry), diffraction correction data has not been developed. However, bond corrections are not necessary as these methods generate the wave directly on the sample surface. Generally, the signal levels are quite small, and require high gain and specialized circuits or devices to bring the signals to usable levels.

6. CORRECTIONS TO PULSE MEASUREMENTS: AN EXAMPLE

With some pulsed systems, and with care in preparing the sample and in taking the measurements, one can expect round-trip time determinations to have standard deviations in the

neighborhood of one hundred picoseconds. As corrections for bond thickness and diffraction are often larger than this, it is necessary to correct for these sources of systematic error where possible. As an example, consider, for a properly prepared sample, that the pulse-echo overlap technique is to be used to measure the round-trip time for a tone burst. The time between the first-received and the second-received echos will be determined. For simplicity, we will treat the case of only one round trip in the sample. The measured time, T , is composed of several terms: the true travel time, δ , the bond thickness contribution, Δ^b , and the diffraction contribution, Δ^D .

a. Bond Thickness

The model for the sample-bonded transducer system has been developed by Williams and Lamb (1958), and has been used extensively by McSkimin (1961) and by Papadakis (1967). They show that by treating the transducer bond sample system as an acoustic transmission line, it is possible to calculate the effects of bond thickness on the measurement of transit time. McSkimin (1961) shows that by measuring the change in the transit time when the system is detuned by a known amount, it is possible to determine the bond thickness in terms of measured quantities. This permits one to correct for the transit time through the bond material. Papadakis (1967) discusses the correction in some detail, and applies it to his measurements on fused quartz and silicon.

Consider an undamped transducer bonded to one end of a sample with some bonding material of known velocity and density; a relatively long wave is reflected from the interface between the sample and the transducer bond system. The reflected wave from this interface experiences a phase shift from the impedance mismatch, which can be calculated from the model. The calculation uses the real and imaginary parts of the effective impedance of the components to obtain the contribution to the phase shift. By detuning the frequency of the applied tone burst, we can measure a corresponding change in transit time. Then, a comparison with the measured values permits the calculation of the bond thickness, so that its effect on the measurement can be determined. The approach is outlined below.

The reflection coefficient for the system described above is given by

$$\frac{E_b}{E_i} = \frac{Z_d - Z_s}{Z_d + Z_s} \quad (42)$$

where E_b is the reflected pressure wave, E_i is the incident pressure wave, Z_s is the specific acoustic impedance of the sample ($Z_s = \rho c$, where ρ is the mass density, and c is the wave propagation speed), and Z_d is the effective specific acoustic impedance of the transducer bond system, given by McSkimin (1961),

$$Z_d = jZ_1 \frac{\left[\frac{Z_1}{Z_2} \tan k_1 \Delta_1 + \tan k_2 \Delta_2 \right]}{\left[\frac{Z_1}{Z_2} - \tan k_2 \Delta_2 \tan k_1 \Delta_1 \right]}, \quad (43)$$

where Z_1 and Z_2 are the specific acoustic impedances of the bond material and transducer material respectively. Using similar conventions, k_1 and k_2 are propagation constants in the respective materials, as Δ_1 and Δ_2 are the respective bond and transducer thicknesses, and j is the $\sqrt{-1}$.

The phase angle (McSkimin, 1961) is calculated by writing Eq. (42) in complex polar form, and gives

$$\gamma = \arctan \left(\frac{2|Z_d|Z_s}{|Z_d|^2 - Z_s^2} \right) \quad (44)$$

where $|Z_d|$ is the modulus of Z_d , and $(\pi + \gamma)$ is the phase shift between the reflected and incident pressure waves impinging on the interface of the transducer bond system with the sample.

When the system is driven at the transducer resonant frequency, f_r can be calculated by using the fact that the transducer at fundamental resonance has a thickness of one half of a wavelength; then $|Z_d|$ becomes $Z_1 \tan k_1 \Delta_1$, which is generally quite small for a thin bond. Frequencies of off-transducer resonance, generally chosen as $0.9 f_r$, give a larger phase angle, since the dependence of Z_d upon off-resonance frequency excitation is large. As outlined in McSkimin (1961) and Papadakis (1967), we can calculate the time difference between phase-matched echo trains, ΔT , caused by the change of drive frequency, and compare it with experimental results to determine the correct condition for overlap.

Assume an alignment of, for example, the first and second echos in the sample. The change in the measured time between echos is given by

$$\Delta T = \frac{1}{f_L} \left(n - \frac{p\gamma_L}{2\pi} \right) - \frac{1}{f_R} \left(n - \frac{p\gamma_R}{2\pi} \right) \quad (45)$$

where p is the number of round trips (for our case, $p = 1$) of the ultrasonic tone burst in the sample, n is the number of cycles of mismatch in the tone burst, f_L is the off-transducer resonance drive frequency, f_R is the drive frequency at transducer resonance, γ_L is the phase angle at the off-resonance frequency (radians), and γ_R is the phase angle (radians) at the transducer resonance. One can calculate ΔT for the case of no cycle mismatch, ($n = 0$) and no bond thickness ($\gamma_R = 0$) from Eqs. (43) through (45). The measured value closest to the calculated value determines the correct experimental cycle for cycle match. Following the experimental identification of the correct match, the measured ΔT can be used to determine the correction for bond thickness by adjusting Δ_1 in Eq. (43) and solving Eq. (44) to bring Eq. (45) into agreement with the measured ΔT values. The phase angle at resonance can be determined, and the travel time correction for bond thickness, Δ^B , is given by

$$\Delta^B = \frac{\gamma_R}{2\pi f_R} \quad , \quad (46)$$

where γ_R and f_R are the phase angle, and drive frequency respectively, both taken at transducer resonance.

b. Diffraction Corrections

Diffraction effects for compressional waves in various crystalline symmetries have been treated by several investigators (Seki et al., 1956; Papadakis, 1963, 1964, 1966) while others have treated isotropic media (Benson and Kiyohara, 1974; Khimunin, 1972; Rogers and Van Buren, 1974). Papadakis (1972) has shown that without diffraction corrections, one can expect errors in travel times as large as $0.25/f_R$. He also discusses (1972) diffraction corrections for the technique of pulse-echo overlap technique. Using the dimensionless quantity

$$S = \frac{z\lambda}{a^2} \quad (47)$$

where S is the Seki parameter, a is the transducer radius, λ is the wavelength in the sample, and z is the distance of propagation, he writes the phase correction due to diffraction, Δ^D as

$$\Delta^D = \frac{[\phi(S_n) - \phi(S_m)]}{2\pi f}, \quad (48)$$

where ϕ is the phase shift in radians of the received wave front that is due to diffraction effects.

For the example, the sample is of length L which gives, $S_n = 2L\lambda/a^2$, and $S_m = 4L\lambda/a^2$.

The true travel time, δ , can be written in terms of the measured time, T , and the corrections,

$$\delta = T + \Delta^b + \Delta^D. \quad (49)$$

The thickness of the sample, L , is measured by conventional means, such as a high precision micrometer, and the velocity of sound is given by

$$v = \frac{2L}{\delta}. \quad (50)$$

7. OPTICAL TECHNIQUES

Optical techniques have been used with success in the measurement of sound velocities in transparent media. The methods include diffraction and scattering of light by sound waves. A direct method of optically measuring an acoustic wavelength in the material can also be used.

The diffraction techniques considered here permit the determination of the wavelength of sound. The wavelength, λ^* , is combined with the ultrasonic frequency, f , of the sound beam to calculate its velocity, according to the expression,

$$f\lambda^* = v, \quad (51)$$

where v is the speed of sound in the sample. In any optical technique, however, one must specify the type of diffraction experienced by the light.

a. Fraunhofer Diffraction of Light by Sound

Consider an ultrasonic transducer, bonded to a transparent solid, generating ultrasonic compressional waves that propagate into the solid. Further suppose that the sound wave encounters light traversing the same medium so that the light is diffracted. There are two different physical regimes that can produce the diffraction effects. One involves the formation of a corrugation in the phase fronts of the light that is due to the spatial variation in the index of refraction of the solid, as caused by the sound waves. This is called Raman-Nath diffraction (Raman and Nath, 1935, 1935a, 1936, 1936a, 1936b; Born and Wolf, 1970). The second regime involves the reflection of light from the evenly spaced crests of the sound waves. These reflections occur under some conditions that are similar to x-ray diffraction by a crystal lattice; this is called Bragg diffraction (Bhatia and Noble, 1953). Both types of diffraction can be used to determine the sound velocity of the material.

In order to determine which type of diffraction effect predominates (Nomoto 1942), a dimensionless parameter (Klein et al., 1965), Q , is defined as

$$Q = \frac{K^*{}^2 L}{\mu_0 K} , \quad (52)$$

where K^* is the ultrasonic propagation constant, L is the width of the ultrasonic beam, μ_0 is the index of refraction, and K is the propagation constant of light in vacuum. If $Q > 9$, one has Bragg diffraction. If $Q < 1$, Raman-Nath diffraction occurs. For $1 < Q < 9$, the diffraction is mixed. For illustrative purposes, consider a typical transparent solid ($c_{\text{sound}} \approx 5 \times 10^3$ m/sec, index of refraction ≈ 1.5 , and ultrasonic beam width $\approx 1.27 \times 10^{-2}$ m), illuminated with light at wavelength 632.8 nm. If the ultrasonic frequency is approximately 27 MHz or less, the interaction satisfies Raman-Nath diffraction conditions. If the frequency is greater than approximately 81 MHz, the interaction is governed by Bragg diffraction conditions.

b. Raman-Nath Diffraction

For the case of Raman-Nath diffraction, we consider the light beam impinging on the medium at an angle, ϕ , with the light normal to the sound beam. The location of the diffraction orders are given by the expression

$$\sin(\theta_n + \phi) - \sin(\phi) = \frac{n\lambda}{\lambda^*} \quad (53)$$

where ϕ is the angle of incidence, θ is the angle of diffraction, λ is the wavelength of light in the medium, λ^* is the ultrasonic wavelength, and n is an integer. One can experimentally set $\phi = 0$, which reduces Eq. (53) to

$$\sin(\theta_n) = \frac{n\lambda}{\lambda^*} \quad (54)$$

The ultrasonic wavelength can be determined from Eq. (53). With the measurement of the ultrasonic frequency, one can determine the sound velocity of the compressional wave (Barnes and Hiedemann, 1957). In determining the wavelength measurement and its uncertainty, it is necessary to make an analysis of the optical setup, including the effects of Snell's Law at interfaces.

c. Bragg Diffraction

For the case of Bragg diffraction, the angle of incidence, ϕ , is set to the angle of diffraction, and

$$n\lambda = 2\lambda^* \sin \phi_B \quad (55)$$

where ϕ_B is the Bragg angle. Bragg diffraction is used to measure wave velocities in the frequency range from approximately 100 MHz to the low end of the gigahertz scale. For example (Krischer, 1968), the technique has been used to measure wave velocity to an estimated accuracy of better than 0.1%. It is also useful in measuring the local velocities within a sample (Simondet

et al., 1976; Michard and Perrin, 1978). Measurements with precision of better than 0.01% in homogeneous samples have been reported (Simondet et al., 1976).

d. Direct Measurements (Fresnel Diffraction)

Consider a standing ultrasonic wave in a sample through which collimated light is passed so that the collimated light beam is perpendicular to the sound beam. A measuring microscope or similar optical device is focused so that the image of the wave can be viewed at the instrument focal plane. Because the images of the wave fronts are $\lambda/2$ apart, it is possible to determine the ultrasonic velocity from measurements of ultrasonic frequency and measurements of the wavelength in the medium. The technique is sensitive enough to detect local variations in velocity greater than 0.01% (Hiedemann and Hoesch, 1934, 1937; Mayer and Hiedemann, 1958, 1959).

e. Brillouin Scattering

Consider the scattering of a photon by a high frequency phonon traveling in a specific direction within a crystal. Application of conservation of energy and momentum to the scattering, coupled with the approximation that any photon frequency shift is small (Benedek and Fritsch, 1966; Beyer and Letcher, 1969, pg. 47-50), gives

$$\nu^* = \pm 2\nu \left(\frac{v}{C} \right) \sin \left(\frac{\theta}{2} \right), \quad (56)$$

where ν is the photon frequency, ν^* is the phonon frequency as well as the difference in frequency of the scattered photons, C is the speed of the photon in the medium, v is the speed of the phonon, and θ is the scattering angle. If desirable, the Bragg condition can be used for constructive reinforcement by adjusting θ (Pollard, 1965). The technique gives three lines in the scattered photon spectra, that can be separated and measured with appropriate optical devices, such as Fabry-Perot interferometers, to obtain ν^* . Eq. (56) can be solved for ν in terms of the other quantities. Brillouin scattering is useful in the investigation of the sound velocity of a material near a phase transition (Fleury, 1970, p. 37-42). The technique has been used for both longitudi-

nal and mixed modes. However, in one study on cubic crystals, no Brillouin scattering was observed from an acoustic branch which consisted of pure transverse waves (Benedek and Fritsch, 1966). Uncertainties in the index of refraction, the scattering angle, and the width of the Stokes and anti-Stokes lines influence the accuracy of the determination. The precision is considerably better, however, with values of 0.1% as mentioned.

In stimulated Brillouin scattering, the photon scattering process is dependent upon the intensity of the radiation striking the surface. With high enough light intensity, nonlinear effects occur, which result in scattering by frequencies and harmonics created by harmonic generation (Breuer, 1965). A large buildup of acoustic intensity, both compressional and shear waves, accompanies a threshold in optical intensity (Chiao et al., 1964). Other effects include the possibility of sample destruction from the intense radiation, and the line pulling effects of the laser cavity on the scattered light, which affects accuracy (Fleury, 1970, p. 57-58). Amplified acoustic frequencies have been reported as high as 60 GHz.

III. THEORETICAL AND EXPERIMENTAL RESULTS

8. AB INITIO THEORY

It is clear from Section 2 that the calculation of elastic constants requires an accurate computation of the variation in the total energy of a solid, from equilibrium to distorted configurations. Thanks to the availability of powerful computers, great advances in *ab initio* total-energy calculation have been made in recent years. In this section we briefly discuss the approximations used in *ab initio* theory and summarize the calculated results.

The Hamiltonian of a solid consists of five parts: the two kinetic energies of the ions and electrons, K_I and K_e respectively, and the potential energies U_{II} among ions, U_{ee} among electrons, and U_{eI} between electrons and ions. The Born-Oppenheimer (1927) adiabatic approximation is a simple way to separate the electronic from the ionic variables. In this approximation, because the ion speed is at least two orders smaller than the electron speed, one freezes the ionic motion in a configuration specified by a set of ion position vectors \vec{R}_n , then solves the Schrodinger equation for that part of the Hamiltonian that involves the set of electronic coordinates $\{\vec{r}_i\}$:

$$H \psi_\gamma(\{\vec{r}_i\}, \{\vec{R}_n\}) = E_\gamma(\{\vec{R}_n\}) \psi_\gamma(\{\vec{r}_i\}, \{\vec{R}_n\}) \quad , \quad (57)$$

where $H = K_e + U_{ee} + U_{eI}$. Thus the energy E_γ is a function of the ionic configuration. The lowest energy curve of the sum of $E_g = E_\gamma + U_{II}$ as a function of $\{\vec{R}_n\}$ then serves as the potential energy for the ionic motion. A Taylor expansion of E_g about its minimum value E_0 , i.e., the quantum mechanical analog to Eq. (17), takes the form

$$E_g(\vec{R}_n) = E_0 + \frac{1}{2} \sum_{nm\alpha\beta} \Phi_{nm}^{\alpha\beta} u_{n\alpha} u_{m\beta} + \dots \quad , \quad (58)$$

where $u_{n\alpha}$ is the component of the small displacement $\vec{R}_n - \vec{R}_n(0)$ with $\{\vec{R}_n(0)\}$ the equilibrium ionic positions at the minimum energy E_0 . The force functions $\Phi_{nm}^{\alpha\beta}$ are the second derivatives of E_g with respect to the displacements evaluated at the ion equilibrium positions. The force functions are directly related to the elastic constants. For example, for a Bravais lattice they are given by Aschcroft and Mermin (1976) as

$$c_{\alpha\beta\mu\nu} = -\frac{1}{2} \sum_L L_\alpha \Phi^{\beta\mu}(\vec{L}) L_\nu \quad , \quad (59)$$

where the sum is over all the lattice vectors \vec{L} .

Eq. (57) is computationally the most difficult part of the problem, because it deals with about 10^{23} electrons that are interacting with each other and with the ions; the wave functions have to be the properly antisymmetrized many-body functions. Self-consistent density-functional theory (SCDFT) (Hohenberg and Khon, 1964; Kohn and Sham, 1965; Callaway and March, 1984), which casts the Hamiltonian into a density functional, reduces the many-body problem to an effective one-electron problem. This theory has been tested in many crystalline solids and the resulting elastic constants have been excellent, especially for semiconductors.

In SCDFT, the ground state energy of a solid is completely specified by single-particle wave functions of the occupied states $\{\phi_v\}$. First the electron density is given by

$$\rho(\vec{r}) = \sum_v |\phi_v(\vec{r})|^2 \quad . \quad (60)$$

Then the ground state energy is computed as follows:

$$E_g = K_e + U_{ee} + U_{el} + U_{II} + U_{xc} \quad , \quad (61)$$

where the different terms are given by the expressions

$$K_e = \sum_v \int \phi_v^*(\vec{r}) \frac{p^2}{2m} \phi_v(\vec{r}) d^3r \quad , \quad (62)$$

$$U_{ee} = \frac{e^2}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3r d^3r' \quad , \quad (63)$$

$$U_d = -e^2 \sum_n Z_n \int \frac{\rho(\vec{r})}{|\vec{r} - \vec{R}_n|} d^3r , \quad (64)$$

$$U_{II} = \frac{e^2}{2} \sum'_{mn} Z_m Z_n / |\vec{R}_m - \vec{R}_n| , \quad (65)$$

$$U_{xc} = \int \rho(\vec{r}) \epsilon_{xc}[\rho(\vec{r})] d^3r . \quad (66)$$

Note that $n = m$ is excluded in the sum in Eq. (65). The meaning and the notations of the above equations are mostly self-evident, except the $\epsilon_{xc}[\rho]$ in Eq. (66); this is the correction term arising from the many-body exchange and correlation effects. The square bracket means ϵ_{xc} is a functional of ρ . Several different expressions for ϵ_{xc} as a function of ρ , available in the literature (for example, Wigner, 1934; Hedin and Lunquist, 1971; Ceperley and Adler, 1980; Perdew and Zunger, 1981), have yielded similar results for the structural properties of semiconductors. A minimization of E_g with respect to ϕ_v^* with the constraint that the total number of electrons is a constant, leads to the familiar Schroedinger single-particle equation:

$$\left[\frac{\rho^2}{2m} + V(\vec{r}) \right] \phi_v(\vec{r}) = \epsilon_v \phi_v(\vec{r}) , \quad (67)$$

where the ϵ_v is a Lagrange multiplier and V is an effective one-electron potential containing three parts:

$$V = V_{ee} + V_d + V_{xc} . \quad (68)$$

$$V_{ee} = e^2 \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3r' . \quad (69)$$

V_{eI} is the Coulomb potential due to ionic charges Z_n

$$V_d = -e^2 \sum_n Z_n / |\vec{r} - \vec{R}_n| . \quad (70)$$

Finally, the exchange-correlation potential V_{xc} is given by

$$V_{xc} = \epsilon_{xc} + \rho \frac{\partial \epsilon_{xc}}{\partial \rho} . \quad (71)$$

Thus Eqs. (59) and (67) through (70) form a repeated loop, $\rho \rightarrow V \rightarrow \phi \rightarrow \rho$, and the calculation must be iterated until self-consistency is achieved. Following this recipe for calculating the total energy, one calculates E_g as a function of ionic positions, finds the equilibrium configuration, then imposes a strain and calculates the strain energy to deduce the elastic constants, following the prescription of Section 4. The problem then becomes strictly computational. The most challenging task is an accurate solution for the single-particle eigen states of Eq. (67). For a crystalline solid, lattice translational symmetry simplifies the problem, and band-structure techniques can be applied to obtain the solution. Because the strain energy is many orders of magnitude smaller than the total energy, very precise computation is required if one hopes to obtain meaningful elastic constants. So far at least two band structure methods have been demonstrated as reliable for all three elastic constants: the plane-wave method using pseudopotentials (PP-PW) (Nielsen and Martin, 1983, 1985a), and the full-potential linearized-muffin-tin-orbital method (FP-LMTO) (Methfessel et al., 1989). Although the full-potential augmented plane-wave method (FP-APW) (Krahauser et al., 1979; Wimmer et al., 1981; Wei and Krahauser, 1985; Ferreira et al., 1989) has produced excellent lattice constants, structural energies, and bulk moduli, the complete semiconductor elastic constants based on this method are not yet available.

Even if the total energies at different distortions can be calculated accurately, there is still the problem of searching for the equilibrium atomic positions in a distorted crystal, and the numerical determination of elastic constants from energy differences. If the strain energy can be calculated directly without taking the difference between two large energies, or if the derivatives can be calculated directly, not only can the computation time be shortened but the numerical errors will also be reduced. The quantum mechanical theory of forces and stresses of Nielsen and Martin (1985b), and the closely related direct calculation of elastic constants from linear

response theory of Baroni et al. (1987) represent the status of efforts in this direction. The former has been carried out for all three elastic constants for Si, Ge, and GaAs (Nielsen and Martin, 1985a), while the latter has been done only on the bulk modulus for Si; both are based on the PW-PP method. Table 1 shows a comparison between theoretical calculations and experimental results. From this comparison, it is fair to say that we have a very reliable *ab initio* theory for the elastic constants for crystalline semiconductors based on the self-consistent local density-functional theory. Note that Table 1 also lists the results from an empirical tight-binding (TB) theory to be discussed in Section 10.

9. VALENCE FORCE FIELD MODEL

The preceding section showed that *ab initio* theory for the elastic constants requires complicated computations. Accurate *ab initio* calculations for semiconductors have been obtained only recently and only for several systems. On the other hand, phenomenological microscopic models of elastic constants for all semiconductors have been available for some time. Of these, the valence force-field model (VFF) is perhaps the simplest and the most useful. This topic has been reviewed and well analyzed in a paper by Martin (1970), and its conclusions constitute the main body of this section.

a. Diamond Structure

The original VFF model by Musgrave and Pople (1962) was for the diamond structure, in which the elastic energy is a quadratic form, in terms of the changes in each bond length Δr_i , in bond angles $\Delta \theta_{ij}$, and in the products $\Delta r_i \Delta r_j$ and $\Delta \theta_{ij}$, between nearest neighbor bonds. For elastic constants, Keating (1966) showed that the VFF can be simplified by the following approximation to the elastic energy of the crystal:

$$\Delta E = \frac{3\alpha}{8d^2} \sum_i [\Delta (\vec{r}_i \cdot \vec{r}_i)]^2 + \frac{3\beta}{8d^2} \sum_{i>j} [\Delta (\vec{r}_i \cdot \vec{r}_j)]^2 \quad (72)$$

where the bond index i runs over all the bonds, but the i and j sum only over those pairs of bonds that are connected to a common atom. In Eq. (72) d is the equilibrium bond length and $\Delta(\vec{r}_i \cdot \vec{r}_j)$ is the change in the dot product of the two bond vectors which start at the common atom, point along the bond directions, and end at the first-neighbor atoms. Following the calculational procedure described in Section 2 for uniform expansion, the U in the VFF under a uniform expansion can be shown to be $U = 2(3\alpha + \beta) e^2 d^2 / \Omega$, where $\Omega = a^3/4$ is the equilibrium volume per unit cell. Thus

$$B = (C_{11} + 2C_{12})/3 = (\alpha + \beta/3)/a \quad . \quad (73)$$

For the shear strain described by Eq. (9), Eq. (72) yields $U = 4\beta e^2/a$. Thus according to Eq. (10),

$$C_{11} - C_{12} = 4\beta/a \quad . \quad (74)$$

For the shear strain described by Eq. (11) and with an internal displacement $u = (0,0,u)$, Eq. (72) yields the following expression,

$$U = [\alpha(e - \eta)^2 + \beta(e + \eta)^2] / (8a) \quad , \quad (75)$$

where $u = \eta a/4$. A comparison between Eqs. (13) and (74) shows that $\phi = 16(\alpha + \beta)/a^3$, $D = -(\alpha - \beta)/a^2$ and $C_{44}^{(0)} = (\alpha + \beta)/a$. Using these results in Eq. (14), we find that the Kleimann

internal displacement parameter in the present model is given by

$$\zeta = (\alpha - \beta) / (\alpha + \beta) = 2C_{12} / (C_{11} + C_{12}) \quad . \quad (76)$$

Eq. (15) then produces

$$C_{44} = 2\alpha\beta / [(\alpha + \beta)a] \quad . \quad (77)$$

The three elastic constants given above are not independent, and can be shown to relate to each other by the Keating Identity (1966), or

$$I_k = 2C_{44}(C_{11} + C_{12}) / [(C_{11} - C_{12})(C_{11} + 3C_{12})] = 1 \quad . \quad (78)$$

b. Zincblende Structure and Coulomb Force

The Keating Identity (1966) holds very well for systems with the diamond structure but not so well for the zincblende compounds (see Table 3). One obvious difference between the two structures is the presence of Coulomb interactions arising from charge shifts between the cation and anion sublattices in zb semiconductors. Martin (1970) incorporated Blackman's (1959) treatment of the Coulomb forces in the Keating VFF in the following manner. First, the Coulomb energy was treated as a screened Madelung energy E_M . For example, in a uniformly expanded crystal with a bond length r , the Coulomb energy was taken to be $E_M = -N\alpha_M Z^{*2}e^2/(\epsilon r)$, where N is the total number of unit cells, $\alpha_M = 1.6381$ is the Madelung constant, and Z^{*2}/ϵ is the effective charge defined by the optic-mode splitting:

$$S = Z^{*2}/\epsilon = \mu (\omega_l^2 - \omega_t^2) / (4\pi e^2) . \quad (79)$$

In Eq. (79) ω_l and ω_t are respectively the longitudinal and transverse phonon frequencies in the long-wave-length limit. Then to counterbalance the Coulomb forces, a repulsive force term was added and assumed to contribute to the bond-stretching energy in the form

$$\Delta E_R = - \sum_i \alpha_M Z^{*2} e^2 \Delta r_i / (4\epsilon d^2) . \quad (80)$$

With the above two contributions added, the total strain energy is $\Delta E_T = \Delta E + \Delta E_M + \Delta E_R$, where ΔE is the VFF contribution in Eq. (72) and ΔE_M is the change in the Madelung energy. The energies ΔE_T are expanded in a power series, and only terms up to the second power in the strain are kept. The ΔE_M contributions arising from fixed values of the charge shift $S = Z^{*2}/\epsilon$ on the atomic sites under different strains were worked out by Blackman (1959). Using these results and defining

$$s = e^2/d^4 S = e^2 Z^{*2} / (d^4 \epsilon) , \quad (81)$$

Martin (1970) obtained the following modified expressions for the elastic constants:

$$C_{11} + 2C_{12} = (3\alpha + \beta) / a - 0.355s \quad , \quad (82)$$

$$C_{11} - 2C_{12} = 4\beta/a + 0.053s \quad , \quad (83)$$

$$\zeta = [(\alpha - \beta) / a - 0.294s] / C_M \quad , \quad (84)$$

$$C_{44} = (\alpha + \beta) / a - 0.136s - C_M \zeta^2 \quad , \quad (85)$$

where C_M is defined as

$$C_M = (\alpha + \beta) / a - 0.266s \quad . \quad (86)$$

The above equations can be combined to yield

$$\zeta = (2C_{12} - C') / (C_{11} + C_{12} - C') \quad , \quad (87)$$

where $C' = 0.314s$. Since the extra parameter s is fixed by the optic modes and the bond length, the above results combine into a new identity, the Martin Identity (1970):

$$I_M = \frac{2C_{44} (C_{11} + C_{12} - C')}{(C_{11} - C_{12}) (C_{11} + 3C_{12} - 2C') + 0.831C' (C_{11} + C_{12})} \quad . \quad (88)$$

Table 3 lists a set of experimental values of the elastic constants and the s values for a number of diamond and zincblende semiconductors. These values are used to compute the force constants α and β and the identity expressions I_k and I_M given in Eqs. (77) and (88) respectively. Several sets of data are quoted for some of the systems to show the uncertainties in the experiments for these systems. The table results clearly show that the inclusion of the Coulomb energies improves the identity relation; the deviations of I_M from unity are 15% or less. Also listed are the values for another identity relation, I_{BoM} from Eq. (119), based on a tight-binding model to be discussed in Section 10. Martin further studied trends as functions of the bond lengths d and the ionicity scale f_i of Phillips and Van Vechten (1969). He found that α scales roughly as $1/d^3$, i.e.,

$$\alpha d^3 / e^2 = \text{constant} \quad , \quad (89)$$

where e is the electron charge. He also found the ratio between the bond-angle and bond-stretching forces tend to decrease as f_i increases and scales roughly as

$$\beta/\alpha \propto 1 - f_i \quad (90)$$

He further observed that if S of Eq. (79) is set equal to f_i and if the α and β values are extrapolated, using Eqs. (89) and (90), from those fitted to the average values of the B and $C_{11} - C_{12}$ for Si and Ge, then all the elastic constants can be predicted from Eqs. (82) through (86) to an accuracy of 10%.

It is interesting to compare Eq. (82) using the results of Eqs. (89) and (90), with Cohen's (1985) empirical formula for the bulk modulus

$$\beta = (1971 - 22\lambda) / d^{3.5} \quad , \quad (91)$$

where B is in GPa, d in Å, and $\lambda = 0, 1$, and 2 respectively for the group IV, III-V, and II-VI semiconductors. Both Martin and Cohen's formulas give B values to better than 10% for all materials tabulated in Table 3. The B in Eq. (91) scales as $1/d^{3.5}$, while in VFF it scales as $1/d^5$.

10. TIGHT BINDING THEORY

In the semi-empirical tight-binding (TB) approach, the total energy of a semiconductor crystal is assumed to be the sum of the electron energies $\epsilon_v(\vec{k})$ in the valence bands plus repulsive pair energies u_{ij} between the nearest-neighbor atoms (Chadi, 1978):

$$E_T = E_{bs} + U_r = \sum_v \sum_k \epsilon_v(\vec{k}) + \sum_{i>j} u_{ij} \quad . \quad (92)$$

Furthermore, the band energies are interpolated by using a TB Hamiltonian which contains term values of the atoms, and a handful of interaction parameters between orbitals of the neighboring atoms. Despite the simplicity of Eq. (92), recent first-principles theories have given some support to this approximation.

One virtue of the TB approach over valence forcefield models is that it is a quantum theory without much complication. As compared to first-principles theory, the TB approach is easier to execute, particularly when applied to complicated systems such as alloys and superlattices. In actual applications, the TB calculation either is carried out using the full band-structure calculation (BS), or is approximated by simpler local theory such as Harrison's (1980, 1983a and 1983b) bond orbital model (BOM).

a. Band Structure Calculations

The simplest TB Hamiltonian contains the s and p atomic term values ϵ_s and ϵ_p for both cations and anions, and the nearest-neighbor two-center interactions $V_{ss\sigma}$, $V_{sp\sigma}$, $V_{pp\sigma}$ and $V_{pp\pi}$. To be more explicit, the 8×8 k-dependent Hamiltonian contains the term values as the diagonal matrix elements while the off-diagonal matrix elements between the cation and anion orbitals are given by

$$H_{\gamma\gamma'}(\vec{k}) = \sum_{\vec{d}} e^{i\vec{k} \cdot \vec{d}} h_{\gamma\gamma'}(\vec{d}) , \quad (93)$$

where the sum runs over the four first neighbor atoms specified by the bond displacements \vec{d} . The γ 's are the orbital indices for s, p_x , p_y and p_z . The $h_{\gamma\gamma'}$ values are related to the two-center V's by the Slater-Koster (1954) relations:

$$h_{ss} = V_{ss\sigma} , \quad (94)$$

$$h_{sx} = \alpha_1 V_{sp\sigma} , \quad (95)$$

$$h_{xx} = \alpha_1^2 V_{pp\sigma} + (1 - \alpha_1^2) V_{pp\pi} , \quad (96)$$

$$h_{xy} = \alpha_1 \alpha_2 (V_{pp\sigma} - V_{pp\pi}) , \quad (97)$$

where $\alpha_i = x_i/d$ are the direction cosines of \vec{d} and the V's depend only on the length d.

Once the values of these TB parameters and their dependences on the bond length are known, the Hamiltonian at each k inside the Brillouin zone (BZ) can be evaluated, and the summation of k carried out to obtain the band-structure energy, which, when added to the repulsive energy, gives the total energy of any specified geometry. All the elastic constants, associated internal displacements, and transverse optical phonon frequencies, are readily calculable following the procedure of Section 2. The only point to note is the k -sum, which without strain can be calculated accurately by using the ten special k points (Chadi and Cohen, 1973) in the irreducible wedge of the BZ. Under strain, the crystal symmetry changes; one needs to extend these special k points to other nonequivalent wedges. However, since the sum of the valence-band energies as a function of k is a rather smooth function, a uniform sampling over the whole BZ converges very quickly. A $5 \times 5 \times 5$ grid is sufficiently accurate for the required calculations. To avoid the numerical inaccuracy inherent in direct energy subtractions, one can also calculate the second derivatives directly by using perturbation theory.

Perturbation theory starts with the expansion of the k -dependent Hamiltonian H in powers of the infinitesimal strain parameter e , keeping terms up to second order:

$$H(k) = H_0 + H_1 e + \frac{1}{2} H_2 e^2, \quad (98)$$

where H_0 is the strain-free Hamiltonian, and H_1 and H_2 are respectively the first and second derivatives with respect to e evaluated at $e = 0$. The band energy contribution to the strain coefficient then comes from the second derivative of E_{bs} with respect to e , denoted by

$$\frac{\partial^2 E_{bs}}{\partial e^2} = \sum_v \sum_k \langle v\vec{k} | H_2 | v\vec{k} \rangle + 2 \sum_v \sum_c \sum_k \frac{|\langle v\vec{k} | H_1 | c\vec{k} \rangle|^2}{\epsilon_v(\vec{k}) - \epsilon_c(\vec{k})}, \quad (99)$$

where $\epsilon_c(\vec{k})$ and $|c\vec{k}\rangle$ are respectively the eigen energies and eigenvectors of H_0 for the conduction bands, and $v\vec{k}$ stands for the valence bands. Note that the inter-valence-band contributions in the second-order perturbation sum cancel exactly so they are not needed in Eq. (99). To evaluate

these matrix elements one needs to have the first and the second strain derivatives of the two-center interactions and the direction cosines α_i . For the strain parameters e defined in Section 2, and for the two center interactions V that scale as $1/d^n$, the following results are useful.

(i) For the bulk modulus, the direction cosines do not change, and we have

$$\partial V / \partial e = -nV, \text{ and } \partial^2 V / \partial e^2 = n(n+1) V.$$

(ii) For $C_{11} - C_{12}$ with e specified in Eq. (9), we get

$$\partial \alpha_i / \partial e = \alpha_i (\delta_{i1} - \delta_{i2}), \partial^2 \alpha_i / \partial e^2 = \alpha_i [3(\delta_{i1} + \delta_{i2}) - 4] / 3, \partial V / \partial e = 0, \text{ and } \partial^2 V / \partial e^2 = 4nV/3.$$

(iii) For C_{44} with the strain e given in Eq. (11) and an internal displacement u , we find

$$\begin{aligned} \partial V / \partial e &= -n\alpha_1\alpha_2V, \partial V / \partial u = n\alpha_3v/d, \partial \alpha_i / \partial e = (\delta_{i1}\alpha_2 + \delta_{i2}\alpha_1) / 2 - \alpha_i\alpha_1\alpha_2, \\ \partial \alpha_i / \partial u &= -\delta_{i3}/d + \alpha_i\alpha_3/d, \partial^2 V / \partial e^2 = n(n-1)V/9, \partial^2 V / \partial u^2 = n(n-1)V/(3d^2), \\ \partial^2 \alpha_i / \partial e^2 &= -\alpha_i(\delta_{i1} + \delta_{i2}) / 12, \partial^2 \alpha_i / \partial u^2 = -2\delta_{i3}\alpha_3/d^2, \\ \partial^2 V / (\partial e \partial u) &= -n(n+2)\alpha_1\alpha_2\alpha_3/d, \text{ and finally} \\ \partial^2 \alpha_i / (\partial e \partial u) &= (\delta_{i1}\alpha_2\alpha_3 + \delta_{i2}\alpha_1\alpha_3 + 2\delta_{i3}\alpha_1\alpha_2 - 6\alpha_i\alpha_1\alpha_2\alpha_3) / (2d). \end{aligned}$$

Also note that three second derivatives of the band-structure energy are needed, namely $\partial^2 E_{bs} / \partial e^2$, $\partial^2 E_{bs} / (\partial e \partial u)$, and $\partial^2 E_{bs} / \partial u^2$ for the evaluation of $C_{44}^{(0)}$, D , and ϕ of Eq. (13) respectively.

b. Bond Orbital Model

Harrison's (1980, 1983a and 1983b) bond orbital model (BOM) emphasizes calculations of the TB total energy in terms of local energies. One special feature of the BOM is its universality. Another feature is that its simple, and often analytical forms provide direct insight into the essential physics. Although BOM aims at predicting trends, it is reasonably accurate in many cases.

The band-structure energy, or the center of gravity of the valence band, in BOM is computed in the following steps. The terms involved are indicated in Figure 2.

(1) Construct the sp^3 hybrid orbitals $|h\rangle$ for each atom; these hybrid orbitals are directed toward the neighboring atoms. For example, the hybrid in the $[111]$ direction is given by

$$|h\rangle = (|s\rangle + |p_x\rangle + |p_y\rangle + |p_z\rangle) / 2 \quad (100)$$

The hybrid energy is then given by $\epsilon_h = \langle h | H | h \rangle = (\epsilon_s + 3\epsilon_p) / 4$. In the zinc-blende structure, the cation hybrid energy ϵ_h^C is in general different from the anion hybrid energy ϵ_h^A . Note that the two hybrid orbitals of the same atom but in two different directions are now coupled by the so-called metallic energy $V_1 \equiv \langle h | H | h' \rangle = (\epsilon_s - \epsilon_p) / 4$.

(2) Construct the bonding and antibonding molecular orbitals, $|b\rangle$ and $|a\rangle$, from the two hybrid orbitals $|h^C\rangle$ and $|h^A\rangle$ directed toward each other along the same bond by diagonalizing a 2×2 matrix with ϵ_h^C and ϵ_h^A on the diagonal, and $V_2 = \langle h^C | H | h^A \rangle$ as the off-diagonal matrix elements. The resulting energies for $|b\rangle$ and $|a\rangle$ are $\epsilon_b = \bar{\epsilon}_h - \sqrt{V_2^2 + V_3^2}$ and $\epsilon_a = \bar{\epsilon}_h + \sqrt{V_2^2 + V_3^2}$ respectively, where $\bar{\epsilon}_h = (\epsilon_h^C + \epsilon_h^A) / 2$ is the mean hybrid energy, V_2 is the covalent energy, and $V_3 = (\epsilon_h^C - \epsilon_h^A) / 2$ is the polar energy. The eigen states can also be written explicitly in terms of these energies:

$$|b\rangle = \sqrt{(1 + \alpha_p)/2} |h^A\rangle + \sqrt{(1 - \alpha_p)/2} |h^C\rangle \quad (101)$$

$$|a\rangle = -\sqrt{(1 - \alpha_p)/2} |h^A\rangle + \sqrt{(1 + \alpha_p)/2} |h^C\rangle \quad (102)$$

where α_p is called the polarity and is defined as

$$\alpha_p = V_3 / (V_2^2 + V_3^2)^{1/2} \quad (103)$$

(3) The quantity ϵ_b would be the center of gravity of the valence bands, if interactions between states on different bonds were neglected. Harrison (1983b) incorporated these interactions in a perturbation theory in which the change of the bonding energy is given by

$$\Delta\epsilon_b = \sum_{a'} |\langle b | H | a' \rangle|^2 / (\epsilon_b - \epsilon_{a'}) \quad (104)$$

where the sum runs over the antibonding states of the surrounding bonds. Note that the interactions among the bonding states lead to the formation of the valence bands, but do not shift their center of gravity; therefore these interactions need not to be considered in the total energy calculation. This simplification transpires the major physics in the tetrahedral bonding of semiconductors. Including the energy correction $\Delta\epsilon_b$, the final band-structure energy per bond (which contains two electrons) is given by

$$E_b = 2\epsilon_b + 2\Delta\epsilon_b \quad . \quad (105)$$

Harrison (1980) referred to the second term as the metallization energy. Besides providing a simple means for evaluating the center of gravity of the valence electron, Harrison's BOM also provides a set of universal TB parameters. Based on comparison with the free-electron band width (Froyen and Harrison, 1979) and with empirical TB parameters, Harrison (1983b) deduced the following set of universal two-center interactions:

$$V_{\alpha\alpha'} = \eta_{\alpha\alpha'} \hbar^2 / (md^2) \quad , \quad (106)$$

with $\eta_{ss\sigma} = -1.32$, $\eta_{sp\sigma} = 1.42$, $\eta_{pp\sigma} = 2.22$ and $\eta_{pp\pi} = -0.63$, where m is the free-electron mass, and d the bond length. In units where d is in \AA and V in eV, $V_{\alpha\alpha'} = 7.62 \eta_{\alpha\alpha'} / d^2$. The pair-repulsive energy u in the BOM is taken as resulting from the overlap of wavefunctions of the orbitals on the two centers, and was shown to have the form

$$u = u_0 (d_0/d)^4 \quad , \quad (107)$$

where d_0 is the equilibrium bond length. The value of u_0 is determined by requiring that d_0 is the experimental value. Note that the d dependences of both $V_{\alpha\alpha'}$ and u are taken to be the proper scaling not only among different systems but also within the same system, as the bond length varies under distortions.

Under these assumptions, the bond energy E_{bond} , which is defined as the difference between the energy per bond in a semiconductor and the average energy per two electrons in the

free atoms, i.e., $E_{\text{bond}} = E_b + u_0 - 2\bar{E}$, takes the following simple form for a nonpolar semiconductor (Harrison 1983a):

$$E_{\text{bond}} = V_2 (1 - \alpha_m + 9\alpha_m^2 / 16) , \quad (108)$$

where α_m is called the metallicity and is defined as $2V_1/V_2$. For a polar semiconductor, E_{bond} becomes slightly more complicated,

$$E_{\text{bond}} = 2\bar{E}_h - 2\bar{E} - 2(V_2^2 + V_3^2)^{1/2} \left[1 - \frac{1}{2}\alpha_c^2 + \frac{9}{16}\alpha_c^4 (V_{1C}^2 + V_{1A}^2)/(V_2^2 + V_3^2) \right] , \quad (109)$$

where $\alpha_c = \sqrt{1 - \alpha_p^2}$ is called the covalency, and V_{1C} and V_{1A} are the metallic energies for the cation and the anion respectively. The bulk modulus also takes a very simple form; for a group-IV semiconductor, it reads

$$B = -2V_2 (1 - 9\alpha_m^2/16)/(\sqrt{3} d^3) , \quad (110)$$

and for a polar semiconductor, it becomes

$$B = -2V_2 \left[\alpha_c^2 - \frac{9}{8}\alpha_c^3 (5\alpha_c^2 - 4)(V_{1C}^2 + V_{1A}^2)/(V_2^2 + V_3^2) \right] / (\sqrt{3} d^3) . \quad (111)$$

These expressions show that the bulk modulus varies as $1/d^5$ in the pure covalent case, and as $1/d^9$ in the extreme ionic limit $V_3 \gg V_2$. Note that this result is different from the $1/d^{3.5}$ dependence in Cohen's (1985) formula and the $1/d^4$ scale in VVF.

Shear strains cause a semiconductor to distort away from perfect tetrahedral symmetry. To deal with the shear elastic coefficient, the BOM has to be modified. A simple approximation is the rigid hybrid model (Harrison, 1983b; van Schilfgarde and Sher, 1987) in which the hybrid orbitals of each atom are assumed to remain in their original tetrahedral directions despite the lattice distortion. Then the hybrids of two nearest neighbor atoms making up the bonding and antibonding states no longer are directed toward each other, as shown in Figure 3. There is a misalignment angle θ between each hybrid and the line connecting the two atoms, and the covalent energy V_2 is given by

$$V_2(\theta) = \frac{1}{4} [V_{ss\sigma} - 2\sqrt{3} \cos \theta V_{sp\sigma} - \cos^2 \theta V_{pp\sigma} + 3(1 - \cos^2 \theta) V_{pp\sigma}] . \quad (112)$$

The lowest-order change δV_2 , caused by an infinitesimal angular misalignment $\delta\theta$, is then given by

$$\delta V_2 = \frac{1}{4} (\sqrt{3} V_{sp\sigma} + 3V_{pp\sigma} - 3 V_{pp\sigma}) (\delta\theta)^2 . \quad (113)$$

Under the strain e described in Eq. (9) for $C_{11} - C_{12}$, there is no bond length change, and $(\delta\theta)^2 = 2e^2/3$. If one assumes that the metallization coupling is only through the metallic energies V_{1c} and V_{1A} , as has been assumed so far, then the change of the crystal energy is the change of the band-structure energy due to δV_2 . Then according to Eqs. (10) and (105),

$$\begin{aligned} C_{11} - C_{12} &= \frac{\sqrt{3}}{2d^3} \frac{\partial E_b}{\partial V_2} \delta V_2 / e^2 \\ &= \frac{\sqrt{3}}{4d^3} \alpha_c (\sqrt{3} V_{sp\sigma} + 3V_{pp\sigma} - 3V_{pp\sigma}) \left[1 + \left(\frac{3}{4} - \frac{9}{8} \alpha_c^2 \right) (V_{1c}^2 + V_{1A}^2) / (V_2^2 + V_3^2) \right] \end{aligned} \quad (114)$$

Harrison (1983b) has pointed out, however, that in addition to V_1 , other interactions such as V_1^* shown in Figure 2 produce important contributions to the shear elastic constant. By arguing that these other contributions must cancel those associated with the change δV_2 arising from the metallization energy in a rigid rotation, Harrison deduced the following expression:

$$C_{11} - C_{12} = \frac{\sqrt{3}}{4d^3} \alpha_c^3 (\sqrt{3} V_{sp\sigma} + 3V_{pp\sigma} - 3V_{pp\sigma}) . \quad (115)$$

Under the strain e for the C_{44} given in Eq. (11) and with an internal displacement given by $u = \eta d / \sqrt{3}$ as described in Section 2, the bond misalignment angles for the four bonds have the same magnitude with $(\delta\theta)^2 = 2(\eta + e/2)^2/9$. The four bond lengths also change, with the change for one pair given by $\delta r_1 = \delta r_2 = \delta + \epsilon$ and by $\delta r_3 = \delta r_4 = -\delta + \epsilon$ for the other pair, where $\delta = (e - \eta) d/3$ and $\epsilon = (\eta + e/2)^2 d/9$. If again one assumes that the metallization is only through V_1 , then the strain energy density can be shown to be given by

$$\begin{aligned} U &= 9B\delta^2/(2d^2) + 3(C_{11} - C_{12}) (\delta\theta)^2 \\ &= B(e - \eta)^2/2 + (C_{11} - C_{12}) (\eta + e/2)^2/3 . \end{aligned} \quad (116)$$

For a given strain ϵ , U can be minimized with respect to η , which yields the Kleimann displacement parameter $\zeta = \eta/\epsilon$ with ζ given by

$$\zeta = (B - C/3) / (B + 2C/3) , \quad (117)$$

where $C = C_{11} - C_{12}$. Finally from $U = C_{44}\epsilon^2/2$ the following relationship is established:

$$9/C_{44} = 6/C + 4/B , \quad (118)$$

or

$$I_{BOM} = 9BC / [C_{44} (6B + 4C)] = 1 . \quad (119)$$

If one includes the effect of V_1^* , the energy density will involve an additional term which couples δr and $\delta \theta$. Then the analysis is no longer simple.

The above explicit formulas for the elastic constants in BOM are not much more complicated than the valence-force-field model. However, they relate macroscopic forces to intrinsic atomic interactions. It is interesting to note that the simple identity relation of Eq. (119) holds very well. As can be seen in Table 3, this result is certainly better than the I_k of Eq. (78), and is very competitive to Martin's identity I_M which requires the inclusion of his particular treatment of Coulomb forces. Because the Coulomb energy is not included explicitly in Eq. (118), its contribution to the elastic constants is probably small.

c. Numerical Results and Quantitative Applications

To study the quantitative aspect of the theory, one first needs to establish the TB parameters. Table 4 lists the term values we will use. The values of the outermost valence levels are taken to be minus the experimental first ionization energies listed in Kittel's (1986) book, and the other term values are deduced from calculated extraction or promotion energies using norm-conserved atomic pseudopotentials (Bachelet et al., 1982). These term values are very similar to Mann's (1967) Hartree-Fock calculations used by Harrison (1980), which are also given in

Table 4. The major difference in the two sets occurs in heavy elements, where relativistic s-shifts are important, but were not included in Mann's results.

Table 5 lists the values of bond lengths, bond energies E_{bond} , the elastic coefficients B , $C \equiv C_{11} - C_{12}$, and C_{44} , and the zone-center TO phonon frequencies ω for a selected group of systems to be examined in the remainder of this section. Experimental values are also presented, with the exception of extrapolations for the elastic constants of AlP and AlAs. Table 6 compares results between the BOM and the full band-structure (BS) calculation using Harrison's universal TB parameters. Except for the Ge and HgTe values, the agreement between the two calculations for E_{bond} is within 10% or better. The calculated E_{bond} values are also in fair agreement with the experimental values except for diamond. Since diamond has a much smaller bond length than the rest of the systems, this discrepancy is an indication of a limit to the scaling rules for both $V_{\alpha\alpha'}$ and u . Although the trends for the bulk moduli from both calculations are similar, the calculational errors in the BOM can be as large as 50%. Also note that the calculated values of B for most systems are only about one half of the experimental values. The tabulated values of C and C_{44} for BOM are based on Eqs. (115) and (118) respectively. Considering the simplicity of these formulas, the agreement with the band structure calculations is remarkable. However, the overall calculated values for these two shear coefficients are also consistently smaller than the experimental values.

The above comparisons show that the BOM and the BS calculations predict similar qualitative trends for the binding energies and the elastic constants. In this regard, the BOM has the advantage of providing explicit forms to show the dependences on bond lengths and polarities. However, the merit of the TB theory over the valence-force model is in its ability to incorporate atomic quantities to mimic quantum mechanical effects. To be useful for specific material science applications, the theory has to be more quantitative. Successful quantitative application of the TB theory has been made by Chadi (1978, 1979 and 1984) in his study of semiconductor surfaces. The comparisons in Table 6 indicate that the BOM should be treated differently than the BS calculation when considered for quantitative applications. If one wishes to calculate the

properties using the local picture, one should use the BOM. If one wants to carry out the TB Hamiltonian precisely, one needs to adopt a different set of parameters based on the BS calculation. For this reason, we shall next consider quantitative applications of BOM and BS calculations separately.

There are many ways to parametrize the TB theory. Chadi (1979,1984) used a simple form for the repulsive energy $u = a + b (d - d_0) + c (d - d_0)^2$ and the same $1/d^2$ scaling for the TB parameters $V_{\alpha\alpha'}$. To keep the theory as close to Harrison's (1983a and 1983b) form as possible but free it from the $1/d^2$ and $1/d^4$ scaling rules for $V_{\alpha\alpha'}$ and u respectively, we assume the following forms,

$$V_{\alpha\alpha'} = V_{\alpha\alpha'}^o (d_0/d)^n \quad (120)$$

and

$$u = u_0 (d_0/d)^m \quad (121)$$

where the superscript and subscript o indicate the values evaluated at the equilibrium bond length d_0 . For simplicity, the values of $V_{\alpha\alpha'}^{(o)}$ are taken to be Harrison's values given in Eq. (106) scaled by a factor f :

$$V_{\alpha\alpha'}^{(o)} = f V_{\alpha\alpha'}^{\text{Harr}} \quad (122)$$

Thus there are four parameters for each system: the scaling parameter f , the powers n and m , and the value u_0 . These parameters can be determined by requiring that the model produce the correct experimental values for E_{bond} , d_0 , $C_{11} - C_{12}$, and B . Since $C_{11} - C_{12}$ is only governed by $V_{\alpha\alpha'}^{(o)}$ in both BOM and band calculation, it alone determines the scaling factor f . Then the bond energy E_{bond} can be used to determine u_0 . The requirement that the first derivative of E_T is zero at d_0 then determines the ratio of the powers n/m , which couples with the equation for the bulk modulus to yield the values for n and m . One can then use these sets of parameters to check the validity of the model by calculating other quantities not employed in the fitting, e.g., C_{44} , the internal displacement parameter ζ , and the optical phonon frequencies ω at the zone

center. If the results are acceptable, the model can be extended to more complicated systems such as alloys and superlattices with local environments similar to the bulk crystals.

Table 7 shows the results for f , n , m , and u_0 obtained from the above fitting procedure by using the full band-structure calculations, and the corresponding values of C_{44} , ζ , and ω calculated for consistency checks. The scaling factor f ranges from 1 to 1.4 and tends to decrease with an increase in polarity. In the power dependence of $V_{\alpha\alpha'} \propto (d_0/d)^n$, n ranges from 2.8 to 4.3, which is larger than the $n = 2$ used in Harrison's universal TB parameters. For the repulsive pair energy $u = u_0(d_0/d)^m$, the power m ranges from 3.8 to 6.8. The ratio m/n falls in the range from 1.3 to 1.9, which is smaller than the $m/n = 2$ used by Harrison. The calculated values of C_{44} for most systems agree with the experimental data to 10% or better, except for diamond and ZnS. Note that the experimental data for ZnS are rather dispersed. The calculated TO optical phonon modes at Γ in $1/\text{cm}$ for most group IV and III-V systems also agree with experiments to 10% or better. The discrepancies for the II-VI systems are larger, about 15%. Reliable results for ζ from both experiments and first-principles calculations are available only for a limited number of systems. The calculated ζ in the TB model agrees very well with those results, as shown in Table 1. The overall results for C_{44} , ζ and ω in the TB calculations are equivalent to those based on the valence force model including Martin's Coulomb force corrections. By construction the TB model also produces the correct cohesive energies, bond lengths, bulk moduli and shear coefficients $C_{11} - C_{12}$, because these quantities are used to fit the parameters.

The results in Table 7 are based on the term values given in Table 4 and the TB parameters scaled from Harrison's universal parameters. It is useful to know how the predictions are influenced by these parameters and the fitting procedure. Table 8 shows the results based on Chadi procedure in which the TB matrix elements $V_{\alpha\alpha'}$ are scaled as $1/d^2$, and the repulsive pair energy is taken to be $u = u_0 + u_1 (d - d_0) + u_2 (d - d_0)^2$. The parameter u_0 is set to produce the correct bond energy, u_1 is determined by requiring the correct equilibrium bond length, and u_2 is fixed by the bulk modulus. Two sets of TB parameters are tabulated for each system: one is the set used by Chadi (1978, 1979 and 1984), the other is the set obtained by multiplying Harrison's

$V_{\alpha\alpha'}$ by the scaling factor f listed in Table 7. For convenient comparison, the zero of the term values is set equal to the anion s energy. Despite considerable differences in these two sets of TB parameters, the results of the predictions from both sets are very similar and also very similar to those predicted from the other procedure used for the results in Table 7. The only noticeable difference between the predictions in Table 8 and Table 7 is that the present procedure produces larger phonon frequencies and slightly smaller C_{44} values. We also note that the fitted parameters u_0 , u_1 , and u_2 and the predicted values for Chadi's elastic constant set in Table 8, are not the same as Chadi's published (1979) values; these give bulk moduli about 20% smaller than the experimental values, but give considerably better phonon frequencies.

To parameterize the BOM, several different stages of approximations can be made. However, for a general application, the full BOM steps presented in Eqs. (100) to (104) for calculating E_b should be followed, regardless of approximations. The simplest model, referred to as BOM(1), is to include only V_{1A} and V_{1C} in the the matrix element $\langle b | H | a' \rangle$ for the calculation of the metallization energy in Eq. (104). The next approximation, BOM(2), is to include V_1^x as well. Finally one can include all the first-neighbor interatomic TB parameters in $\langle b | H | a' \rangle$; this will extend the $|a'\rangle$ to those belonging to the second-neighbor bonds. This last approximation will be referred to as BOM(3).

Table 9 shows the results for BOM(1) following the parameterization procedure described in Eqs. (120) to (122). The fitted parameters f , n , m , and u_0 are substantially different from those based on the BS calculations. The predicted C_{44} for the group IV and III-V systems are slightly larger than the experimental values, but good for the II-VI systems. The calculated ζ values, although not all smaller than those in Table 7, are smaller on the average. The predicted phonon frequencies are too high.

The parameters and the predicted results from BOM(2) are listed in Table 10. These parameters more closely resemble those in Table 8 than do the values from BOM(1). However, the predicted C_{44} values are still too small, and the ζ values are too large, but the ω values are better than those from BOM(1).

Table 11 contains the results from BOM(3); these approach those of the BS calculations. In comparison with Table 7 and the experimental values in Table 5, the BOM(3) does well for ω , produces slightly smaller C_{44} , and probably slightly larger ζ values.

In conclusion, the TB method is a reasonable approach to the static elastic properties of semiconductors. If carried out rigorously, the TB parameters in Table 7 will provide quantitative results for superlattices, alloys, and possibly surfaces in which the local environments are similar to those in the bulk. The quantitative predications of BOM are not as good as the BS calculations, but are still reasonable. The fitted parameters given in Tables 9 to 11 allow different stages of approximations to be made using the BOM. This is especially useful for more complicated systems, because computationally the BOM is about two orders faster than the band structure calculations.

11. SEMICONDUCTOR ALLOYS

The systems to be considered in this section are alloys of the diamond and zincblende semiconductors, both the ordered and disordered alloys. The ordered alloys include binary compounds such as SiC, and ternary compounds, such as GaInAs₂, in three crystal structures of the types CuAuI, chalcopyrite, and CuPt as shown in Figure 4 (Bernard et al., 1988). The disordered alloys include binary solutions such as Si_{1-x}Ge_x and pseudobinaries such as Hg_{1-x}Cd_xTe and GaAs_{1-x}Sb_x, where x is the fractional concentration. These alloys have been widely used and studied; however, detailed information about their elastic constants is scarce both experimentally and theoretically. One reason for the lack of rigorous calculation is that the elastic constants of these systems are more complex; existing theories are not as accurate, particularly for disordered alloys. Another reason may be attributed to the fact that most properties of these alloys, including their elasticity, were thought to be reasonably well approximated by the concentration weighted averages of their constituents. Because of the rudimentary state of the theory, we shall deal only with the simplest elastic constant, the bulk modulus. Our focus is on the difference between the alloy bulk modulus B and the concentration weighted averaged value \bar{B} , i.e.,

$\Delta B = B - \bar{B}$. There are several fundamental questions that can be addressed. Is ΔB positive or negative? Does the sign and the magnitude of ΔB depend on the state of order? How can we calculate the bulk modulus in a disordered system? Analysis of these questions constitutes the content of this section.

a. Ordered Alloys

Long-range ordering has been found to exist in many epitaxially grown III-V semiconductor alloys (Kuan et al., 1985; Jen et al., 1986; Ihm et al., 1987; Klem et al., 1987; Huang et al., 1987; Gomyo et al., 1986, 1987, and 1988; Norman et al., 1987; Shahid et al., 1987). Wei and Zunger (1989) have given a rather complete list of these ordered alloys, most of which are of the form ABC_2 existing in three different types of structures, CuAuI, chalcopyrite, and CuPt; all have their atomic planes stacked as ACBCACBC, but along three different directions, (100), (210), and (111) respectively, as shown in Figure 4. There have been several first-principles calculations made to study the structural properties of these compounds; these focused mainly on the cohesive energies and bond lengths. As indicated in the preceding section, one virtue to fitting the TB model for the bulk semiconductors is to use it for interpolating alloy properties. Table 12 lists the results for the bulk moduli of a number of III-V and II-VI alloys derived from full TB band structure calculations using the parameters given in Table 7. Also listed are the average values and percentage deviations from the mean $\Delta B / \bar{B}$. Note that all ΔB values are negative and that most of the magnitudes are small, except for Ga_2AsSb and Ga_2PSb ; the latter has the largest difference in the constituent compounds among the alloys listed. Although the magnitudes of ΔB get larger for systems with larger differences in the bond lengths, the dependence does not seem to be a simple function of the bond length difference. The uniformly negative ΔB values also appeared in the first-principles local density functional calculations for ordered $GaAsSb$ alloys by Ferreira et al. (1989).

The reason for the negative values of ΔB , in a very qualitative argument, is that the bulk moduli of semiconductors scale inversely as high powers of the lattice constant, and, at the same

time, the alloy lattice constant is approximated well by the mean value, by Vegard's law (1921). This implies that the value of B at the mean lattice constant should lie below the straight-line average. Since the TB results for the bulk moduli constants should not be qualitatively different from the valence force field model (VFF) predictions, most of the key physics for the bowing of B should be contained in a VFF analysis. The major effects in the VFF can in turn be realized from the following simple analysis.

Consider the local structure of a CuAuI or chalcopyrite crystal ABC_2 . Focus on a local tetrahedral cluster A_2B_2C with the two A atoms and two B atoms on the vertices of the tetrahedron and the C atom near the center. Let the coordinates of the two A atoms be $(-1, 1, -1)d/\sqrt{3}$ and $(-1, 1, -1)d/\sqrt{3}$ and the two B atoms be at $(1, 1, 1)d/\sqrt{3}$ and $(1, -1, -1)d/\sqrt{3}$. Let the force constants be k_A and k_B , and equilibrium bond lengths be d_A and d_B for the AC and BC bonds respectively. To attain equilibrium, the central C atom is displaced by $(\epsilon, 0, 0)d/\sqrt{3}$. We further define mean values $\bar{d} = (d_A + d_B)/2$ and $\bar{k} = (k_A + k_B)/2$, the relative differences $\delta_o = (d_A - d_B)/\bar{d}$ and $\Delta_o = (k_A - k_B)/\bar{k}$, and $d = \bar{d}(1 + \delta)$. Then the AC bond is stretched by an amount $\bar{d}(\delta + \epsilon/3 - \delta_o/2)$ from its equilibrium value and similarly the BC bond is compressed by $\bar{d}(\delta - \epsilon/3 + \delta_o/2)$. The strain energy for any arbitrary δ and ϵ is given by $\Delta E = \bar{d}^2 [k_A (\delta + \epsilon/3 - \delta_o/2)^2 + k_B (\delta - \epsilon/3 + \delta_o/2)^2]$. When ΔE is minimized with regard to δ and ϵ , one finds $\delta = 0$ and $\epsilon = 3\delta_o/2$, and the minimum ΔE is zero. If the crystal expands uniformly with δ having a fixed small value, then ϵ becomes $\epsilon = 3\delta_o/2 - 3\Delta_o\delta$ and $\Delta E = 2\bar{k}(1 - \Delta_o^2/4)\bar{d}^2\delta^2$. Thus the effective spring constant is

$$k_{\text{eff}} = \bar{k} \left(1 - \Delta_o^2/4\right), \quad (123)$$

which is smaller than the average value \bar{k} . This weakening of the restoring force constant in the alloy is due to the internal displacement, represented by ϵ in the above model, which provides an extra degree of freedom for relaxation in response to the external stress. The bulk modulus B is proportional to k_{eff}/d , so the alloy bulk modulus minus the mean B is then given by

$$\Delta B = \bar{B} \left(\delta_o\Delta_o - \delta_o^2 - \Delta_o^2/4\right), \quad (124)$$

where we recall the definitions $\delta_0 = (d_1 - d_2) / \bar{d}$ and $\Delta_0 = (k_1 - k_2) / \bar{k}$. Since δ_0 and Δ_0 tend to have different signs, the bond length difference gives an extra negative contribution to ΔB (the first two terms). If both the bond-stretching force constant α and the bond-angle restoring force β in the VVF are included, the equilibrium value of ΔE is no longer zero; but the deviation ΔB can also be shown to be similar to Eq. (123) and is given by

$$\Delta B = \bar{B} \left[\delta_0 \left(\frac{3\Delta\alpha + \Delta\beta}{3\alpha + \beta} \right) - \delta_0^2 - 3 \frac{(\Delta\alpha)^2}{(\alpha + 2\beta)(3\alpha + \beta)} \right] / 4, \quad (125)$$

where $\Delta\alpha = (\alpha_1 - \alpha_2)$ and $\alpha = (\alpha_1 + \alpha_2) / 2$ and similarly for $\Delta\beta$ and β . Eq. (125) reduces to (124), if β is set equal to zero.

The above descriptions illustrate two mechanisms for the negative ΔB values, the $1/a^q$ scaling of B , with q ranging from 3.5 to 9, and more degrees of freedom for internal relaxation. Quantitative results should be described by TB, because in addition to the strain energy, there is also some chemical effect built into the TB theory. Although these ordered compounds have been found from epitaxial growth, the bulk moduli are probably difficult to measure, because these alloys are not single bulk crystals and because the ordering is only partial. It is interesting to note that the B values for a SiC/AlN alternating layer superlattice along (100) and for the constituent compounds have been calculated by Lambrecht and Segall (1990) using the LMTO; their percentage deviation from the mean $\Delta B / \bar{B}$ was found to be about -2%, which falls in the range of the ternary alloys in Tables 12 and 13.

Not all the mechanisms considered above apply to the ordered compounds of the elemental semiconductors, because internal relaxation under pressure may not be allowed, e.g., if the structure is assumed to be zincblende. Unfortunately, a simple analysis of the elastic constants of the 4-4 compounds cannot yet be made, because the tight-binding and VVF parameters have not yet been extended to deal with the atomic pairs not existing in the constituent crystals. However, several first-principles calculations have been made on the ordered SiC and SiGe (Martins and Zunger, 1986; Qteish and Resta, 1988; van Schilfgaarde, 1990). The main results are listed in

Table 13; the theoretical results were calculated for the zincblende structure. The plane-wave pseudopotential (PP-PW) calculation of Martins and Zunger (1986) for SiC gave a -21% value for $\Delta B/\bar{B}$, which is in reasonable agreement with the experimental value of -17%. This is consistent with the qualitative argument based on the $1/d^5$ scaling of B . The theory also yielded a negative formation energy, which also agrees with experiment. For SiGe, the theoretical calculations cited in Table 14 gave positive formation energies, which are consistent with the fact that no ordered bulk compounds of SiGe have been grown. However, some weak ordering has been found in the epitaxial SiGe films (Ourmazd and Bean, 1985). The calculated values of $\Delta B/\bar{B}$ for the zincblende SiGe are either slightly above or just below zero. These differences, however, fall within the uncertainties of the present first-principles theory. The best conclusion which can be drawn from these results is that the B value for SiGe should be very close to the mean value.

b. Disordered Alloys

Disordered binary alloys $A_{1-x}B_x$ of diamond semiconductors and pseudobinary alloys $A_{1-x}B_xC$ of zincblende semiconductors AC and BC are considered in this section; they are not amorphous materials as they still possess their constituent diamond and zincblende lattices respectively, as characterized by their crystal diffraction patterns. We shall consider the pseudobinaries first. The alloying atoms A and B in these alloys belong to a fcc sublattice, and the C atoms to the other sublattice. However, the positions of the A and B atoms are not necessarily locked precisely on the lattice sites. The extended x-ray absorption fine-structure spectroscopy (EXAFS) data (Mikkelsen and Boyce, 1982 and 1983; Boyce and Mikkelsen, 1985; Balzarotti et al., 1984 and 1985) have consistently shown a bimodal distribution of the bond lengths in these alloys, although the average lattice constant follows the Vegard (1921) law $a = (1 - x)a_{AC} + xa_{BC}$. Figure 5 shows an example of the results for the bond lengths in $Ga_{1-x}In_xAs$ deduced by Mikkelsen and Boyce (1982) from their EXAFS data. In a first approximation, the crystal structure of an alloy can be viewed as having the A and B atoms on their fcc sublattice with an average lattice constant, while the C atoms are distorted away from their lattice sites, in a way

similar to the local structures of the three ordered superlattices considered earlier. The difference is that there is no long-range superlattice ordering of the A and B atoms in the disordered state. This simple crystal picture is only a first approximation; the above EXAFS experiments also indicate that the sublattice of the A and B atoms is less than a perfect fcc. There are also theoretical calculations (e.g., Sher et al., 1987; Ferreira et al., 1989) which suggest a certain degree of short-range ordering in these alloys, namely that the arrangement of the A and B atoms is not random. Figure 6 shows an example of the calculated deviations (Sher et al., 1987) of the probabilities from random distribution, $\Delta p_n = p_n - p_n^0$ for $\text{Ga}_{1-x}\text{In}_x\text{As}$ as a function of alloy concentration x , where p_n is the probability of having n Ga atoms and $4-n$ In atoms on the vertices of a local tetrahedral cluster in the alloy, and $p_n^0 = {}_4C_n (1-x)^n x^{4-n}$, where ${}_4C_n$ is a binomial coefficient, is the corresponding value for the random distribution.

The structural energy needed for calculating the elastic constants of a disordered alloy is an ensemble average of the total energy over the distribution of the alloying atoms under strains. It has been demonstrated (Ferreira et al., 1989) that the total energy of a semiconductor can be decomposed into the sum of multisite correlation energies, from the single-site, the pair, and up to a cluster containing a handful of sites. In other words, the multisite correlation energies converge to zero quickly at a manageable number of sites. This implies that the structural energy of an alloy is an average of these multisite correlation energies. Connolly and Williams (1983), working on metal alloys, proposed that these multisite correlation energies be deduced from the ordered systems that are composed of the same atoms. This scheme allows a direct application of the first-principles theory in the calculation of the energy parameters. These energetics can then be used in the alloy statistics such as in the cluster variational method of Kikuchi (1951) or in the Monte Carlo calculations to deduce the distribution functions or the average properties. This theory has been carried out extensively for semiconductor alloys by Ferreira et al. (1989), and respectable results have been obtained for the phase diagrams and alloy equilibrium properties. For this theory to fit the elastic constants requires detailed dependences of the multisite energies under different strains that have yet to be worked out. Also, the validity of using the

energy parameters deduced from ordered alloys in the disordered systems needs to be examined further.

There is a different cluster approach, which directly relates the alloy Hamiltonian to the distribution function (Gautier et al., 1975; Chen et al., 1987; Berra et al., 1988; Dreysse et al., 1989). In this approach one focuses on a particular cluster in an alloy ensemble. The average energy per cluster can be written (Chen et al., 1988) as

$$\langle \epsilon \rangle = \sum_n \sum_m \left(\epsilon_n + \frac{1}{2} h_{nm} \right) p_n P_{nm} . \quad (126)$$

where ϵ_n is the energy of a cluster detached from a given alloy configuration, and h_{nm} is the interaction energy, which is the change of energy of the combined system when the cluster is put back into the alloy. In Eq. (126), p_n is the probability that a cluster is of the type n , specified by the number of A and B atoms and their arrangements, and P_{nm} is a conditional probability that the surrounding environment is in state m when the cluster is in state n . The factor $1/2$ in Eq. (126) is to eliminate the double counting in the total average alloy energy $\langle E \rangle = M \langle \epsilon \rangle$, where M is the ratio between the size of the alloy and the cluster. If one writes $\langle \epsilon \rangle = \sum p_n \epsilon(n)$, then an effective cluster energy can be defined as

$$\epsilon(n) = \sum_m \left(\epsilon_n + \frac{1}{2} h_{nm} \right) P_{nm} . \quad (127)$$

This procedure is particularly useful when the interaction energy is short ranged. One can start with a small cluster and a given probability distribution, then calculate the effective cluster energies from Eq. (126). These energies are then used in a statistical theory to deduce a cluster distribution, which in turn is used to calculate a new set of $\epsilon(n)$ and distribution functions, and the process is iterated until it converges. It should be pointed out that to compute the total energy of an alloy quantum mechanically, one needs to solve the Schroedinger equation for a Hamiltonian which does not have the lattice translational symmetry so indispensable in traditional band-structure theory. If the fluctuation of the alloy potential from the virtual crystal approximation (VCA), where the alloy potential is approximated as the concentration weighed

average, is small, then the next leading correction to VCA can be obtained from perturbation theory. This should work for most semiconductor alloys except for systems with large potential fluctuations such as $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (Chen and Sher, 1982; Spicer et al., 1982; Hass et al., 1983). A more general but more difficult approach is to extend the present molecular coherent potential approximation (MCPA) (Hass et al., 1984) to clusters and to achieve a triple self-consistency (Chen et al., 1987): consistency between cluster distribution and Hamiltonian, between the Hamiltonian and electron density, and between the self-energy operator Σ in the cluster CPA theory and the potential fluctuations. To date this theory has been carried out only for metal alloys, and then only within the single-site KKR-CPA with a random distribution (Schwartz and Bansil, 1975; Gyorffy and Stocks, 1978). Major work is needed to determine if this approach can achieve the same degree of rigor for disordered alloys as self-consistent density functional theory, which has been successfully used in dealing with crystalline semiconductors.

The above idea has been applied to an elastic medium model to deduce a mean field theory for the internal strain and bulk modulus in semiconductor alloys (Chen et al., 1988). This theory starts by assuming that the alloy has an effective lattice constant and effective modulus. When part of the effective alloy medium is replaced by a specified cluster, there will be strain energy introduced. It was shown that this strain energy can be taken as the effective energy $\epsilon(n)$ for that cluster; the probability distribution p_n within a statistical theory can then be deduced. The internal strain energy is calculated as $E = M\langle\epsilon\rangle = M\sum p_n\epsilon(n)$. When the alloy is under an external pressure δP , the effective cluster energy will change by an amount $\delta\epsilon(n)$, which implies a change of the total energy by an amount $\delta E = M\langle\delta\epsilon(n)\rangle$. Then the bulk modulus of the alloy can be obtained from $\Delta E = \frac{1}{2}(\delta P)^2 V/B$, where V is the alloy volume. The mean-field nature of this approach is evident from the fact that the calculation requires knowledge of the alloy lattice constant and elastic constants which are only assumed and are required to be calculated self-consistently. To illustrate this self-consistency procedure, let us consider the following simple spring model for a random pseudobinary alloy $A_{1-x}B_xC$. The cluster corresponds to the four bonds surrounding an "impurity" atom A or B, and the environment of the cluster corresponds to

the twelve bonds which connect inwardly to the cluster and outwardly to a rigid lattice of the effective alloy. It is worth mentioning that there have been detailed analyses of the valence-field force models for the strain energies of semiconductor alloys in regard to the range allowed for lattice relaxation (Martins and Zunger, 1984; Chen and Sher, 1985). It was found that by neglecting the bond-angle forces, one can use a shorter range of lattice relaxation to obtain the correct mixing enthalpies and bond lengths. The simple model considered here works amazingly well for these properties.

Let the spring constants for the pure AC and BC compounds be k_A and k_B respectively, and the effective alloy spring constant be k , with similar notations for the bond lengths d_A , d_B and d . When an A atom is embedded in the medium, all sixteen bonds under consideration will relax, and the strain energy is given by

$$\epsilon(A) = \frac{1}{2} k_1 (d - d_A)^2 \quad , \quad (128)$$

where

$$k_1 = 4k_A k / (3k_A + k) \quad . \quad (129)$$

A similar energy $\epsilon(B)$ is obtained, when a B atom is embedded. The effective bond length d is obtained from a minimization of the average cluster energy $E = (1 - x) \epsilon(A) + x \epsilon(B)$ with respect to d , which yields

$$d = [(1 - x) k_1 d_A + x k_2 d_B] / [(1 - x) k_1 + x k_2] \quad . \quad (130)$$

When the alloy is compressed, the alloy bond length is reduced to $d(1 - e)$, where e is a macroscopic strain corresponding to the external pressure. The pressure induced strain energy for the sixteen bonds in the medium is $\delta E = 8k(de)^2$, and $\delta E = 2k(de)^2$ for each cluster. Embedding an A atom in this compressed medium, one finds the total strain energy for the sixteen bonds to be

$$E_A = \frac{1}{2} k_1 (d - d_A - 4de)^2 \quad . \quad (131)$$

To obtain the extra cluster energy $\delta\epsilon(A)$ induced by the pressure, we subtract $\epsilon(A)$ of Eq. (127) and the background energy for the surrounding twelve bonds from E_A to give

$$\delta\epsilon(A) = -4k_1 (d - d_A) de + 8k_1 (de)^2 - 6k (de)^2 \quad (132)$$

Similarly the following expression for $\delta\epsilon(B)$ is obtained when the embedded atom is B:

$$\delta\epsilon(B) = -4k_2 (d - d_B) de + 8k_2 (de)^2 - 6k (de)^2 \quad (133)$$

Thus the change of the average cluster energy that is due to the pressure is given by $\delta E = \langle \delta\epsilon(n) \rangle = (1 - x) \delta\epsilon(A) + x \delta\epsilon(B)$, which, when equated to $2k(de)^2$, leads to the following self-consistent equation for the effective spring constant k : $k = (1 - x) k_1 + x k_2$. The k can now be solved analytically when both the expression for k_1 in Eq. (129) and the similar expression for k_2 are used. The result is

$$k = \langle k \rangle \left[1 - 3x(1 - x) (\delta k / \langle k \rangle)^2 \right] \quad (134)$$

where $\langle k \rangle = (1 - x) k_A + x k_B$ is the mean spring constant and $\delta k = k_A - k_B$ the difference. It is interesting to compare this result for the 50/50 alloy, i.e., $k = \bar{k} \left(1 - \frac{3}{4} \Delta_o^2 \right)$ with the value $\bar{k} \left(1 - \Delta_o^2 / 4 \right)$ in Eq. (123) for the ordered alloys in the CuAuI and chalcopyrite structures. The alloy spring constant is slightly below the straight line average, and the bowing is larger for a disordered alloy than for the corresponding ordered compound. Using the effective spring constant of Eq. (134), we find that the effective bond lengths for most alloys also bow slightly below their mean value,

$$d = \langle d \rangle + 4x(1 - x) (d_A - d_B) (k_A - k_B) k / [(3k_A + k)(3k_B + k)] \quad (135)$$

because the spring constant tends to increase as the bond length decreases.

To compare the calculations above with experimental data for pseudobinary alloys, we were able to find results for GaAlAs (Landolt-Bornstein, 1988), CdZnTe, CdMnTe, and HgCdTe (Quadri et al., 1986). For GaAlAs, the following linear x dependences were measured (Landolt-Bornstein, 1988): $C_{11} = 11.85 + 0.14x$, $C_{12} = 5.38 + 0.32x$, and $C_{44} = 5.94 - 0.05x$. This lack of

detectable bowing is expected, because of the nearly equal bond lengths of the two constituent compounds and small differences in the elastic constants. The bulk moduli in the three II-VI alloy systems mentioned were obtained from high-pressure x-ray diffraction data. For HgCdTe, the results are similar to that of the GaAlAs in that both the bond lengths and the bulk moduli of HgTe and CdTe are so close that the differences of B between the alloys and the pure crystals were beyond the experimental resolution. However, a 5% Zn in CdZnTe alloy was found to give a 15% increase in the B value from the pure CdTe value and a 10% Mn in CdMnTe gave a 21% decrease (Quadri et al., 1986). These significantly large changes in the B values caused by smaller concentrations cannot be explained from the above considerations.

The qualitative model considered above does not apply to the binary alloys $A_{1-x}B_x$, because in these alloys both the A and B atoms can be found in both sublattices, and the local bond length arrangement is more complicated than the pseudobinary alloys. However, one can expect that there are still more degrees of relaxation in the disordered binaries than in the ordered compounds. Therefore one would conclude that the bulk modulus of the disordered 50-50 SiGe alloy would have a smaller value than those tabulated in Table 13 for the ordered compounds. At least one would not expect the alloy B values to be significantly larger than the mean values \bar{B} . However, the only experimental data available (Bublik et al., 1974), Table 15, show that all three elastic constants for these alloys at three different concentrations exceed the values for Si, and that the $\Delta B / \bar{B}$ is as large as 20%, despite the fact that the bond length difference between Si and Ge is only about 3% and the measured alloy lattice constants are only bowed slightly below the average. This, and the unexplained results for the II-VI alloys, point to the need for a more systematic study of the elastic properties of semiconductor alloys, both experimentally and theoretically.

IV. DISLOCATIONS AND HARDNESS⁷

Hardness has proven to be a useful probe of the mechanical properties of the brittle semiconductors. Here we will use the term hardness to refer specifically to Vickers' hardness, unless otherwise noted. In the Vickers' hardness measurement, a square pyramidal indenter is used and the hardness number is given by the applied load divided by the area of the indentation (i.e., units of pressure). Hardness has been found to be an intrinsic property of the material, because it is relatively independent of the applied load. One advantage of hardness measurements for semiconductors is that, unlike bending tests, only small samples are necessary for conventional Vickers' hardness measurements, or for nanoindenter measurements (Fang et al., 1990), and relatively thin epitaxial films can be probed. Additionally, unlike conventional tests used to measure yield stress, hardness measurements can be made at room temperature, which is far below the usual plastic regime for most semiconductors. As such, the hardness measurement provides a convenient and usable probe.

The question remains, though, as to the interpretation of the hardness measurement in semiconductors: just what property or properties of a semiconductor are we measuring when we measure hardness. In metals, an empirical relationship is found between the hardness, H , and yield stress, Y , such that $H = 3 Y$.

In metals, this relationship can be justified on the basis of continuum theory, as discussed in McClintock and Argon (1966). In semiconductors such a simple relationship between H and σ is not necessarily appropriate for several reasons. During deformation in metals, many slip planes can be active because the Peierls barriers for dislocation motion in most directions are low. In contrast, because the bonds in semiconductors are strongly covalent, the Peierls barriers are high and dislocation in the $(111) \frac{1}{2} \langle 110 \rangle$ slip system dominate.

To date, there is no complete quantitative theory of hardness in the semiconductors in which the temperature dependence, photoplastic effect, and the alloy hardening effect are included. Sher et al. (1985) proposed a model of hardness for the semiconductor compound that gives good quantitative agreement with experiment, but this model does not provide an explanation for several of the observed dependences of hardness. This model of hardness in semiconductors differs from more conventional interpretations and suggests that hardness is dominated by dislocation-dislocation interactions, as opposed to dislocation activation and motion terms. We discuss the results of an improved quantitative model of hardness below.

12. SLIP SYSTEMS

In the dislocation interaction hardness model, Vickers' hardness is found to be dominated by the interaction energy of an idealized array of dislocations that has been generated by the indenter. The idealized array can be considered as a first approximation to the more realistic dislocation tangles found experimentally, leading in higher order to an expansion in dislocation configurations. In this idealized array, no account is taken of the true slip systems active in the semiconductors. Experiments (Hirsch et al., 1985) have demonstrated that, in Vickers' hardness, slip occurs primarily on the $\{111\} \frac{1}{2} \langle \bar{1}10 \rangle$ glide set, where the threefold symmetry of slip and rosette lines occurs at the intersection of the $\{111\}$ planes with the (111) surface.

For indentation on the (111) plane, dislocations can glide on the (111) plane parallel to the surface or on one of the three other $\{111\}$ planes with a total of four active slip planes. Although the detailed analysis differs from that given previously (Sher et al., 1985), the contribution to the hardness from the interaction energy is comparable to that previously calculated. This contribution to H is directly proportional to the shear coefficient.

13. PEIERLS ENERGY

The Peierls energy is difficult to calculate precisely because of dislocation charge effects and reconstruction at the dislocation core. In the context of the hardness measurement we calculate the Peierls energy in order to evaluate the importance of this contribution to the Vickers' hardness number. Although it is generally agreed that dislocations in semiconductors move through the generation and propagation of double kinks, in the hardness measurement, the region about the indenter is grossly plastically deformed. Because the dislocation velocity is low at room temperature (see below), the large dislocation pile-up model proposed by Sher et al. (1985) may be appropriate. If the dislocation separation is small, dislocation motion through kink processes will be suppressed, and the dislocations will propagate as a complete unit.

To get from Configuration A to Configuration B in Figure 7, we must break a row of bonds. Since the long-range strain fields should be comparable in the two configurations as well as in intermediate configurations, the Peierls force can be calculated from local energy considerations only. The energy to break a bond at the dislocation core is approximately given by:

$$U_b = 2\sqrt{V_2^2 + V_3^2} + 2\epsilon_{\text{met}} - V_0 + \frac{n}{4}(\epsilon_h^a - \epsilon_h^c) \quad , \quad (136)$$

where V_2 is the covalent energy, V_3 the ionic energy, ϵ_{met} the metallization energy, V_0 the bond overlap energy, and ϵ_h^a and ϵ_h^c the hybrid energy for the anion and cation, respectively. The first two terms in Eq. (136) account for the loss of the bonding energy of the two electrons in the breaking bond, the third term accounts for regaining the repulsive interaction energy of the bond, and the fourth term accounts for the energy gain to transfer electrons back from the cation to the anion. We note that the electron orbitals of the atoms at the dislocation core are left in the sp^3 hybrids after the bond breaking. The expression in Eq. (136) represents a theoretical maximum of the Peierls energy, since no reconstruction at the core has been included.

To calculate the Peierls energy per unit length, we consider a primary dislocation in the $\langle 110 \rangle$ direction in a zincblende compound. The number of bonds per unit length in $\langle 110 \rangle$ is given by $1/b$ where b is Burger's vector. Thus, the Peierls energy per unit length is given by

$$E_p = \frac{U_b}{b} = \sqrt{\frac{3}{8}} \frac{U_b}{d} , \quad (137)$$

where d is the bond length.

We can now calculate the Peierls force, or the force per unit length necessary to move a dislocation over the potential barrier, as illustrated in Figure 8. The Peierls energy is related to the Peierls force through

$$E_p = F_p L , \quad (138)$$

where

$$L = \frac{b}{2} \quad (139)$$

is the distance between Configurations A and Configuration B. Solving for F_p in terms of U_b and d , we arrive at:

$$F_p = \frac{3}{4} \frac{U_b}{d^2} , \quad (140)$$

or

$$\tau_p = \sqrt{\frac{3}{2}} \frac{3}{8} \frac{U_b}{d^3} . \quad (141)$$

Values for U_b and τ_p are summarized in Table 16.

Now we incorporate the Peierls energy into the hardness model for low temperature where the full barrier must be surmounted. The Sher model is based on energy considerations. The Vickers' hardness number is given by the applied force divided by the area of indentation.

Multiplying the numerator and denominator by h , the depth of indentation, we have $H = E/(W^2h)$, where $E = Fh$ is the energy of indentation, and h is the depth of the indentation.

Including the interaction energy only we have:

$$H = H_{int} = \frac{G \cot \theta}{6\pi (1 - \nu)} \left[-\Delta n \left(\frac{\cot \theta}{2} \right) + \frac{4}{3} + \sin^2 \frac{\theta}{2} \right] \quad (142)$$

where θ is one half the indenter angle. To include the Peierls energy, we consider the total energy necessary to move the dislocations from their initial to final positions in the idealized model. The hardness is then given by:

$$H = H_{int} + H_p \quad (143)$$

where

$$H_p = \frac{E_p}{W^2h} \quad (144)$$

is the Peierls contribution to the hardness, and E_p is the total Peierls energy expended. The total length of dislocation to be moved is calculated to be:

$$L_T = \frac{1}{6} \frac{W^3}{b^2} \cos^2 \theta \quad (145)$$

The total Peierls energy is given by:

$$E_p = \frac{U_b}{b} L_T = \frac{1}{6} \frac{U_b W^3}{b^3} \cos^2 \theta \quad (146)$$

Thus, we have:

$$H_p = \frac{1}{3} \frac{U_b}{b^3} \cos \theta \sin \theta \quad (147)$$

For $\theta = 45^\circ$:

$$H_p = \frac{1}{6} \frac{U_b}{b^3} \quad (148)$$

Values of H_p are summarized in Table 16. Several features of H_p should be noted. First, we have used a zero temperature value of the Peierls energy. Because hardness measurements are typically done at room temperature, one should take the thermal energy into account; this will reduce the values of H_p from those listed in Table 16. Also shown in Table 16 are H_{int} , $H_{int} + H_p$, the best theoretical estimate for H , and H_{exp} . Note that, like H_{int} , H_p is independent of the applied load, in agreement with experiment. Also note that H_p improves the agreement between theory and experiment for the hard, nonpolar materials. For the softer, more ionic materials, H is overestimated by the theory. The overestimation of H may be because of neglect of dislocation velocity effects and their temperature dependence, as discussed below.

14. TEMPERATURE DEPENDENCE

Here we summarize the experimental results and discuss a tentative theory of the temperature dependence of the hardness.

Several recent studies on the temperature dependence of hardness serve to illustrate the behavior. Results for GaAs and Ge are shown in Figure 9. The (111) and (100) faces of GaAs have been examined by Hirsch et al. (1985) and Guruswamy et al. (1986) respectively. Results for the Knoop hardness on the (100) face of n-type Ge are also shown (Roberts et al., 1986). The (100) face of GaAs and the Ge show a definite temperature dependence with a relatively temperature independent region for $T < 450K$ and an exponential temperature dependence for $T > 550 K$:

$$H = H_0 e^{\frac{U}{kT}} \quad (149)$$

with

$$U \cong 0.24 \text{ eV} \quad (150)$$

for (100) GaAs. The results for GaAs (111) appear to follow a similar behavior.

The temperature dependence of hardness suggests that two different mechanisms may determine hardness in the two temperature regimes. At low temperature, the hardness is nearly independent of temperature and may be limited primarily by dislocation interactions. Dislocation mobility is low at low temperatures, and the tendency for dislocation pile-up is high. At elevated temperature, the dislocation mobility is increased, so that dislocations move more readily under an applied stress. Therefore, at higher temperatures, dislocation pile-up is reduced and the hardness is limited by lattice friction, which shows a strong temperature dependence. We are currently investigating a quantitative theory of the temperature dependence of hardness, which also includes the dissimilar velocities of the α and β dislocations.

V. CONCLUDING REMARKS

The experimental methods available to measure elastic constants vary greatly in their accuracy and in the size of samples required. Generally, those that measure the velocity of sound are quite accurate, some yielding elastic constants to one part in 10^6 . These methods are also capable of measuring higher order elastic constants, a subject not treated in this paper. However, the samples required for these measurements must be large, of the order of several centimeters, and must be perfect bulk single crystals. Many semiconductors, alloys in particular, are only grown as thin films on disparate substrates. For these examples, the velocity of sound methods fail, and the less accurate Raman and Brillouin scattering techniques become the methods of choice. Their accuracy is about 1 to 4%, which is adequate for many practical applications.

Most group IV, III-V compound, and II-VI compound semiconductors have been studied, and their elastic constants tabulated. A few remain to be examined, and several should be reexamined because different experimenters do not agree on the results. The situation in the pseudobinary alloys is quite different. Few alloy systems have been adequately studied; those studied have mostly fallen into the class of materials in which the bond lengths of the constituents nearly match. More interesting results are expected from alloys with a bond length mismatch. Such studies would yield a wealth of information on mechanisms responsible for correlations in these alloys, and perhaps even on those responsible for producing the ordered alloys that have been grown recently.

We have emphasized the utility of various parameterized models for treating nonideal situations. However, the most powerful new theoretical developments are in the area of first-principles theories. The advent of self-consistent local density theories (Hohnberg et al., 1964) over

twenty years ago, and the advances in methods to solve the Schrodinger equation, are making real inroads (Methfessel and Van Schilfgaarde, 1990).

Just a few years ago, it took many hours of supercomputer time to solve problems, so only a few existence proofs had been done to demonstrate the accuracy of local density analysis. The new LMTO solution methods (Methfessel and Van Schilfgaarde, 1990; Harris, 1985; Foulkes and Haydock, 1989; Van Schilfgaarde et al., 1991), and the Carr and Parnello extensions (1985) of APW, increase solution speeds into the realm where it is practical to attack many mechanical property problems. Moreover, LDA has now been extended to include many-body corrections (Hybertsen and Louie, 1987), so its few percent deviations from experiments directed at properties sensitive to total energies should soon be corrected. Conduction bands also are corrected by the addition of the many-body effects, so optical and transport properties can also be predicted with accuracy. Surfaces, disordered alloys, and other situations where symmetry is lost, continue to be a challenge, but molecular dynamic techniques, Green's function methods, and variants of the coherent potential approximation, show promise of providing adequate solutions. As computational speeds continue to increase, these methods will evolve into practical engineering tools.

REFERENCES

- Anderson, O.K., Jepson, O., and Glötzl, D. (1985). "Highlights of Condensed Matter Theory,"
edited by F. Bassiani et al., North Holland, Amsterdam.
- Ashcroft, N.W., and Mermin, N.D. (1976). "Solid State Physics," Saunders College,
Philadelphia, 445.
- Bachelet, G.B., Hamman, D.R., and Schluter, M. (1982). Phys. Rev. B26, 4199.
- Balzarotti, A., Czyzk, M.T., Kisel, A., Motta, N., Podgorny, M., and Zimnal-Starnawska, M.
(1985). Phys. Rev. B31, 7526.
- Barnes, J.M., and Hiedemann, E.A. (1957). J. Acoust. Soc. Am. 28, 1218-1221; 29, 865.
- Benedek, G.B., and Fritsch, K. (1966). Phys. Rev. 149, 647-662.
- Bensen, G.C., and Kiyohara, O. (1974). J. Acoust. Soc. Am. 55, 1184-185.
- Berding, M.A., Chen, A.-B., and Sher, A. (1987a). Mat. Res. Soc. Symp. Proc. 90, 127.
- Berding, M.A., Sher, A., and Chen, A.-B. (1987b), J. Vac. Sci. Tech. A5, 3009.
- Berding, M.A., Sher, A., and Chen, A.-B. (J. Appl. Phys., in press).
- Berding, M.A., Sher, A., Chen, A.-B., and Miller, W.E. (1988). J. Appl. Phys. 63, 107.
- Berera, A., Dreysse, H., Wille, L.T., and de Fontaine, D. (1988). Phys. Rev. F18, 149.

- Bernard, J.E., Ferreira, L.G., Wei, S.-W., and Zunger, A. (1988). *Phys. Rev.* B38, 6338.
- Beyer, R.T., and Letcher, S.V., (1969). "Physical Ultrasonics," Vol. 32 in *Pure and Applied Physics*, Consulting Editors H.S.W. Massey and K. A. Brueckner, Academic Press, New York, 47-50.
- Bhatia, A.B., and Noble, W.J. (1953). *Proc. Roy. Soc. London, Ser. A* 220, 356-368, 369-385.
- Blackmann, M. (1959). *Phil. Mag.* 3, 831.
- Bloomfield, P.E., Ferren, R.A., Radice, P.F., Stefanou, H., and Sprout, O.S. (1978). *Nav. Res. Rev.* (May), 1-15.
- Bolef, D.I., deKlerk, J., and Gosser, R.B. (1962). *Rev. Sci. Instrum.* 33, 631.
- Bolef, D.I., and Miller, J.G. (1971). "Physical Acoustics," edited by W.P. Mason and R.N. Thurston, Academic Press, Vol. VIII, 96-201.
- Born, M., and Oppenheimer, J.R. (1927). *Ann. Phys.* 4, 84, 457.
- Born, M., and Wolf, E. (1970). "Principles of Optics," 4th ed., Pergamon Press, Oxford, 593-610. A concise treatment of this subject appears in Chapter 12.
- Boroni, S., Giannozzi, P., and Testa, A. (1987). *Phys. Rev. Lett.* 58, 1861.
- Boyce, J.B., and Mikkelsen, J.C. Jr. (1985). *Phys. Rev.* B31, 6903.
- Breazeale, M.A., and Ford, J. (1965). *J. Appl. Phys.* 36, 3486. For the pure longitudinal mode directions and for isotropic materials, the nonlinear wave equation has been solved here.

- Breazeale, M.A., Cantrell, J.H. Jr., and Heyman, J.S. (1981). "Methods of Experimental Physics, Vol. 19, Ultrasonics," Academic Press, New York, 67-135.
- Brewer, R. (1965). Appl. Phys. Lett. 8, 165.
- Brewer, R.G., and Rieckhoff, K.E. (1964). Phys. Rev. Lett. 13, 334-336.]
- Brugger, K. (1964). Phys. Rev. 133, A1611.
- Bublik, V.T., Gorelik, S.S., Zaitsev, A.A., and Polyakov, A.Y. (1974). Phys. Stat. Soli. B 66, 427.
- Callaway, J., and March, N.H. (1984). Solid State Phys. 38, 135.
- Cantrell, J.H. Jr. (1982). J. Testing and Eval. 10, 223.
- Cantrell, J.H., and Breazeale, M.A. (1977). J. Acoust. Soc. Am. 61, 403.
- Car, M., and Parrinello, M. (1985). Phys. Rev. Lett. 55, 2471.
- Ceperly, D.M., and Alder, B.J. (1980). Phys. Rev. Lett. 45, 566.
- Chadi, D.J. (1978). Phys. Rev. Lett. 41, 1062.
- Chadi, D.J. (1979). Phys. Rev. B19, 2074.
- Chadi, D.J. (1984). Phys. Rev. B29, 785.
- Chadi, D.J., and Cohen, M.L. (1973). Phys. Rev. B8, 5747.
- Chen, A.-B., and Sher, A. (1982). J. Vac. Sci. & Techno. 21(1), 138.
- Chen, A.-B., and Sher, A. (1985). Phys. Rev. B32, 3695.

- Chen, A.-B., van Schilfgaarde, M., Krishnamurthy, S., Berding, M.A., and Sher, A. (1987). In "Ternary and Multinary Compounds," edited by S.K. Deb and A. Zunger, Materials Research Society, Pittsburgh, PA.
- Chen, A.-B., Berding, M.A., and Sher, A. (1988). Phys. Rev. B37, 6285.
- Chern, E.J., Cantrell, J.H., and Heyman, J.S. (1981). J. Appl. Phys. 52, 3200-3204.
- Chiao, R.Y., Townes, C.H., and Stoicheff, B.P. (1964). Phys. Rev. Lett. 12, 592-595.
- Cohen, M.L. (1985). Phys. Rev. B32, 7988.
- Connolly, J.W.D., and Williams, A.R. (1983). Phys. Rev. B27, 5169.
- Conradi, M.S., Miller, J.G., and Heyman, J.S. (1974). Rev. Sci. Instrum. 45, 358-360.
- Cousins, C.S.G., Gerward, L., Staun, L.J., Selsmark, B., and Sheldon, B.J. (1987). J. Phys. C20, 29.
- Dreysse, H., Berera, A., Wille, L.T., and de Fontaine, D. (1989). Phys. Rev. B39, 2442.
- Elmore, W.C., and Heald, M.A., (1969). "The Physics of Waves," McGraw-Hill, New York, 98-104.
- Fang, S., Farthing, L.J., Tang, M.-F.-S., and Stevenson, D.A. (1990). J. Vac. Sci. Technol. A8 (2), 1120.
- Ferreira, L.G., Wei, S.-H., and Zunger, A. (1989). Phys. Rev. B40, 3197.

- Fleury, P.A. (1970). in "Physical Acoustics," Vol. 6, edited by W.P. Mason and R.N. Thurston, Academic Press, New York, N.Y. Chap. 1.
- Foulkes, M., and Haydock, R. (1989). Phys. Rev. B39, 12520.
- Froyen, S., and Harrison, W.A. (1979). Phys. Rev. B20, 2420.
- Gautier, G., Ducatelle, F., and Giner, J. (1975). Phil. Mag. 31, 1373.
- Goldstein, H. (1965). "Classical Mechanics," Addison-Wesley, Reading, MA, 347 ff.
- Gomyo, A., Suzuki, T., and Iijima, S. (1988). Phys. Rev. Lett. 60, 2645.
- Green, R.E. Jr. (1973). "Treatise on Materials Science and Technology, Vol. 3, Ultrasonic Investigation of Mechanical Properties," Academic Press, New York, 7-8, 11-25.
- Guruswamy, S., Hirth, J.P., Faber, K.T. (1968). J. Appl. Phys. 60, 4136.
- Gyorffy, B., and Stocks, G.M. (1978). Inst. Phys. Conf. Ser. 39, 394.
- Harris, J. (1985). Phys. Rev. B31, 1770-1790.
- Harrison, W.A. (1980). "Electronic Structure and the Properties of Solids," Freeman and Co., San Francisco, CA.
- Harrison, W.A. (1983a). Phys. Rev. B27, 3592.
- Harrison, W.A. (1983b). "The Bonding Properties of Semiconductors," SRI International, Menlo Park, CA.
- Hass, K.C., Ehrenreich, H., and Velicky, B. (1983). Phys. Rev. B27, 1088.

- Hedin, L., and Lunquist, B.I. (1971). *J. Phys.* C4, 2064.
- Hiedemann, E.A., and Hoesch, K.H. (1934). *Z. Phys.* 90, 322-326.
- Hiedemann, E.A., and Hoesch, K.H. (1937). *Z. Phys.* 107, 463-473.
- Heyman, J.S. (1976). *Proc. IEEE Ultrason. Symp., Annapolis, MD*, 113.
- Hirsch, P.B., Pirouz, P., Roberts, S.G., and Warren, P.D. (1985). *Philo. Mag.* B52, 759.
- Hohnberg, P., and Kohn, W. (1964). *Phys. Rev.* 136, B864.
- Huang, D., Chyi, J., Klem, J., and Morkoc, H. (1988). *J. Appl. Phys.* 63, 5859.
- Hybertsen, M.S., and Louie, S.G. (1987). *Phys. Rev. Lett.* 58, 1551.
- Ihm, Y.E., Ouska, N., Klem, K., and Morkoc, H. (1987). *Appl. Phys. Lett.* 51, 2013.
- Jen, H.R., Cherng, M.J., and Stringfellow, G.B. (1986). *Appl. Phys. Lett.* 48, 1603.
- Johnson, G.C., and Mase, G.D. (1984). *J. Acoust. Soc. Am.* 75, 1741.
- Keating, P.N. (1966). *Phys. Rev.* 145, 637.
- Khimunin, A.S. (1972). *Acustica* 27, 173-181.
- Kikuchi, R. (1951). *Phys. Rev.* 81, 988.
- Kittel, C. (1986). "Introduction to Solid-State Physics," Table 5, John Wiley & Sons, Inc., New York.
- Kleiman, L. (1962). *Phys. Rev.* 128, 2614.

Klein, W.R., Cook, B.D., and Mayer, W.G. (1965). *Acustica* **15**, 67-74. This work is a theoretical analysis and defines Q.

Klem, J., Huang, D., Morkoc, H., Ihm, Y.E., and Otuska, N. (1987). *Appl. Phys. Lett.* **50**, 1364.

Krahauer, H., Posternak, M., and Freeman, A.J. (1979). *Phys. Rev. Lett.* **43**, 1885.

Krischer, C. (1968). *Appl. Phys. Lett.* **13**, 310.

Kuan, T.S., Kuech, T.F., Wang, W.I., and Wilkie, E.L. (1985). *Phys. Rev. Lett.* **54**, 201.

Kuan, T.S., Wang, W.I., and Wilkie, E.L. (1987). *Appl. Phys. Lett.* **51**, 51.

Lambrecht, W.R.L., and Segall, B. (1990). Private communication.

Landolt-Bornstein (1982). "Numerical Data and Functional Relationships in Science and Technology," Edited by K.-H. Hellwidge, Vol 17, Springer-Verlag, Berlin.

Landolt-Bornstein (1988). "Numerical Data and Functional Relationships in Science and Technology," Edited by K.-H. Hellwidge, Vol. 22, Springer-Verlag, Berlin.

Landau, L.D., and Lifshitz, E.M. (1959). "Theory of Elasticity," Pergamon Press, New York.

Landau, L.D., and Lifshitz, E.M. (1986). "Theory of Elasticity," Third ed., Pergamon Press,

New York. Equation 16 of text is in slightly different form than given on p. 2. Using

$x_i = a_i + u_i$, and substituting into Eq. (16), one obtains the strain tensor found in this reference.

This reference stresses physics concepts to derive the important equations.

Love, A.E.H. (1944). "A Treatise on the Mathematical Theory of Elasticity," Dover, New York.

- Mann, J.B. (1967). "Atomic Structure Calculations, 1: Hartree-Fock Energy Results for Elements from Hydrogen to Lawrencium," Clearinghouse for Technical Information, Springfield, VA.
- Martin, R.M. (1970). Phys. Rev. B1, 4005.
- Martins, J.L., and Zunger, A. (1984). Phys. Rev. B30, 6217.
- Martins, J.L., and Zunger, A. (1986). Phys. Rev. Lett. 56, 1400.
- Mayer, W.G., and Hiedemann, E.A. (1958). J. Acoust. Soc. Am. 30, 756-760.
- Mayer, W.G., and Hiedemann, E.A. (1959). Acta Crystallogr. 12, 1.
- McClintock, F.A., and Argon, A.S. (1966). "Mechanical Behavior of Materials," Addison-Wesley, Reading, MA.
- McSkimin, H.J. (1950). J. Acoust. Soc. Am. 22, 413-418.
- McSkimin, H.J. (1957). IRE Trans. Ultrason. Eng. 5, 25.
- McSkimin, H.J. (1961). J. Acoust. Soc. Am. 33, 12.
- McSkimin, H.J., and Andreatch, P. (1964). J. Appl. Phys. 35, 3312-3319.
- Methfessel, M., Rodriguez, C.O., and Andersen, O.K. (1989). Phys. Rev. B40, 2009.
- Methfessel, M., and van Schilfgaarde, M. (1990). In preparation for Phys. Rev.
- Michard, F., and Perrin, B. (1978). J. Acoust. Soc. Am. 64, 1447-1456.
- Mikkelsen, J.C. Jr., and Boyce, J.B. (1982). Phys. Rev. Lett. 49, 1412.

- Mikkelsen, J.C. Jr., and Boyce, J.B. (1983). *Phys. Rev.* B28, 7130.
- Miller, J.G. (1973). *J. Acoust. Soc. Am.* 53, 710-713.
- Min, B.I., Massida, S., and Freeman, A.J. (1988). *Phys. Rev.* B38, 1970.
- Mitra, S.S., and Massa, N.E. (1982). In "Handbook on Semiconductors," Vol. 1, edited by T.S. Moss, North-Holland, Amsterdam, Chap. 3.
- Murnaghan, F.D., (1951). "Finite Deformation of an Elastic Solid," Wiley, New York. This reference uses matrices throughout and is mathematically more formal.
- Musgrave, J.P., and Pople, J.A. (1962). *Proc. Roy. Soc. (London)* A268, 474.
- Nielsen, O.H., and Martin, R.M. (1983). *Phys. Rev. Lett.* 50, 697.
- Nielsen, O.H., and Martin, R.M. (1985a). *Phys. Rev.* B32, 3792.
- Nielsen, O.H., and Martin, R.M. (1985b). *Phys. Rev.* B32, 3780.
- Nomoto, O. (1942). *Proc. Phys. Math. Soc. Japan*, 24, 380-400, 613-639. This work covers an experimental investigation into the overlap region.
- Norman, A.G., Mallard, R.E., Murgatroyd, I.J., Brooker, G.R., More, A.H., and Scott, M.D. (1987). In "Microscopy of Semiconductor Materials," *Inst. Phys. Conf. Ser.* 87, 77.
- O'Donnell, M., Busse, L.J., and Miller, J.G. (1981). In "Methods of Experimental Physics, Vol. 19, Ultrasonics," Academic Press, New York, 29-65.
- Ourmazd, A., and Bean, J.C. (1985). *Phys. Rev. Lett.* 55, 765.

- Papadakis, E.P. (1963). *J. Acoust. Soc. Am.* **35**, 490-494.
- Papadakis, E.P. (1964a). *J. Appl. Phys.* **35**, 1474.
- Papadakis, E.P. (1964b). *J. Acoust. Soc. Am.* **36**, 414-422.
- Papadakis, E.P. (1966). *J. Acoust. Soc. Am.* **40**, 863-876.
- Papadakis, E.P. (1967). *J. Acoust. Soc. Am.* **42**, 1045.
- Papadakis, E.P. (1972). *J. Acoust. Soc. Am.* **52**, 843.
- Papadakis, E.P. (1976). In "Physical Acoustics," edited by W.P. Mason and R.N. Thurston, Vol. XII, Academic Press, New York, 277-374.
- Patrick, R.S., Chen, A.-B., and Sher, A. (1987). *Phys. Rev. B* **36**, 6585.
- Patrick, R.S., Chen, A.-B., Sher, A., and Berding, M.A. (1988). *J. Vac. Sci. and Technol.* **6**, 2643.
- Phillips, J.C. (1973). "Bonds and Bands in Semiconductors," Academic Press, New York and London.
- Pollard, H.F. (1977). "Sound Waves in Solids," Pion Limited, London, 308-309.
- Prosser, W.H., and Green, R.E. Jr. (1985). In "Proceedings of the 1985 SEM Spring Conference on Experimental Mechanics," Society for Experimental Mechanics, 340-346.
- Qteish, A., and Resta, R. (1988). *Phys. Rev. B* **37**, 1308.

Quadri, S.B., Skelton, E.F., and Webb, A.W. (1986). *J. Vac. Sci. and Technol.* A4(4), 1971, 1974.

Raman, C.V., and Nath, N.S.N. (1935a). *Proc. Indian Acad. Sci. Sect. A, Part I*, 2, 406-412.

Raman, C.V., and Nath, N.S.N. (1935b). *Proc. Indian Acad. Sci. Sect. A, Part II*, 2, 413-420.

Raman, C.V., and Nath, N.S.N. (1936a). *Proc. Indian Acad. Sci. Sect. A, Part III*, 3, 75-84.

Raman, C.V., and Nath, N.S.N. (1936b). *Proc. Indian Acad. Sci. Sect. A, Part IV*, 3, 119-125.

Raman, C.V., and Nath, N.S.N. (1936c). *Proc. Indian Acad. Sci. Sect. A, Part V*, 3., 459-465.

Roberts, S.G., Warren, P.D., and Hirsch, P.B. (1986). *J. Mater. Res.* 1, 162.

Rogers, P.H., and Van Buren, A.L. (1974). *J. Acoust. Soc. Am.* 55, 724-728.

Schwartz, L., and Bansil, A. (1975). *Phys. Rev.* B10, 3261.

Scott, W.R., and Bloomfield, P.E. (1981). *Ferroelectrics* 32, 79-83. Lead attachment to KYNAR films is covered in this paper.

Seki, H., Granato, A., and Truell, R. (1956). *J. Acoust. Soc. Am.* 28, 230-238.

Shahid, M.A., Mahajan, S., Laughlin, D.E., and Cox, H.M. (1987). *Phys. Rev. Lett.* 58, 2567.

Sher, A., Chen, A.-B., and Spicer, W.E. (1985). *Appl. Phys. Lett.* 46, 54.

Sher, A., Berding, M.A., Krishnamurthy, S., van Schilfgaarde, M., Chen, A.-B., and Chen, W. (1987). *Mat. Res. Soc. Symp. Proc.* 90, 91.

- Sher, A., Chen, A.-B., van Schilfgaarde, M., and Berding, M.A. (1987). Proc. 1st Internat. SAMPE Electronics Conf. (23-25 June), Sunnyvale, CA, 323.
- Sher, A., van Schilfgaarde, M., Chen, A.-B., and Chen, W. (1987). Phys. Rev. B36, 4279.
- Sher, A., Berding, M.A., van Schilfgaarde, M., Chen, A.-B., and Patrick, R.S. (1988). J. Crystal Growth 86, 15.
- Simondet, F., Michard, F., and Toquet, R. (1976). Opt. Commun. 16, 411-416.
- Slater, J.C., and Koster, G.F. (1954). Phys. Rev. 94, 1498.
- Spicer, W.E., Silberman, J.A., Morgan, P., Lindau, I., Wilson, J.A., Chen, A.-B., and Sher, A. (1982). Phys. Rev. Lett. 47, 948.
- Stringfellow, G.B., and Greene, P.E. (1969). J. Phys. Chem. Solids 30, 1779.
- Truell, R., Elbaum, C., and Chick, B.B. (1969). "Ultrasonic Methods in Solid State Physics," Academic Press, New York and London.
- Tu, L.Y., Brennan, J.N., and Sauer, J.A. (1955). J. Acoust. Soc. Am 27, 550.
- Van Vechten, J.A. (1969). Phys. Rev. 187, 1007.
- van Schilfgaarde, M. (1990), unpublished.
- van Schilfgaarde, M., Chen, A.-B., and Sher, A. (1986). Phys. Rev. Lett. 57, 1149.
- van Schilfgaarde, M., Chen, A.-B., and Sher, A. (1987). Proc. 7th Internat. Conf. on Ternary and Multinary Compounds (10-12 September 1986), Snowmass, CO, 41.

- van Schilfgaarde, M., Methfessel, M., and Paxton, A.T. (in preparation for Phys. Rev. Lett.).
- van Schilfgaarde, M., and Sher, A. (1987). Phys. Rev. B36, 4375.
- van Schilfgaarde, M., and Sher, A. (1987). Appl. Phys. Lett. 51, 175.
- Vasile, C.R., and Thompson, R.B. (1977). In IEEE Proc. 1977 Ultrasonics Symp., Editors
J. deKlerk and B.R. McAvoy, 84.
- Vegard, L. (1921). Z. Phys. 5, 17.
- Voight, W. (1928). "Lehrbuch der Kristallphysik," Teubner, Leipzig. This contraction of the
indices is also covered in Landau and Lifshitz (1986) and Green (1975), page 6. The
subscripts are contracted as $11 \rightarrow 1$, $22 \rightarrow 2$, $33 \rightarrow 3$, $23 \rightarrow 4$, $31 \rightarrow 5$, $12 \rightarrow 6$.
- Wei, S.-H., and Krahauer, H. (1985). Phys. Rev. Lett. 55, 1200.
- Wei, S.-W., and Zunger, A. (1989). Phys. Rev. B39, 3279.
- Wigner, E.P. (1934). Phys. Rev. 46, 1002.
- Williams, A.O. Jr., (1970). J. Acoust. Soc. Am. 48, 285 (1970).
- Williams, J., and Lamb, J. (1958). J. Acoust. Soc. Am. 30, 308-313.
- Wimmer, E., Krahauer, H., Weinert, M., and Freeman, A.J. (1981). Phys. Rev. B24, 864.
- Wood, D.M., Wei, S.-W., and Zunger, A. (1988). Phys. Rev. B37, 1342.
- Zallen, R. (1982). In "Handbook on Semiconductors," Vol. 1, edited by T.S. Moss, North-
Holland, Amsterdam, Chap. 1.

FOOTNOTES

¹We assume that the angles defining the orientation of axes in the crystal are not specified. For a further discussion, see Landau and Lifshitz (1986).

²The accuracy depends upon the correct choice for the resonant frequency of the transducer, which can be obtained from the fact that it is $1/2$ wavelength in thickness. Using a micrometer, and the wave velocity of the transducer material, one can calculate the resonant frequency of the transducer. Transducer off-resonant conditions can have a relatively large influence on the measurement of round-trip time for pulsed studies.

³Other influences on the measurement of round-trip time are bond thickness and diffraction. These are discussed in various papers, including McSkimin (1961) and Papadakis (1967, 1972).

⁴Information on piezoelectric transducers is available from Valpey-Fisher Corp., 75 South Street, Hopkinton, MA 01748, and Crystal Technology Inc., 1060 E. Meadow Circle, Palo Alto, CA 94303.

⁵A film, sold under the tradename KYNAR is an example of this. Information on this material can be obtained from Pennwalt Corp., Box C, King of Prussia, PA 19406-0018. Some of the properties are covered in Bloomfield et al. (1978). Lead attachment to KYNAR films is covered in Scott and Bloomfield (1981).

⁶For example, suppliers of damped ultrasonic transducers can supply material suitable for high temperature and shear measurements. Two companies are Panametrics, Inc., 221 Crescent St., Waltham, MA 02254, and Harisonics, Inc., 7 Hyde St., Stamford, CT 06907.

⁷Much of this section is adapted from "Final Report" (AFOSR-F49620-85-0023) by M.A. Berding (1988), SRI International, Menlo Park, CA.

FIGURE CAPTIONS

1. Distortion of a tetrahedron corresponding to the C_{44} elastic constant; u is the internal displacement between the anion and cation sublattices.
2. Labels of the interactions between hybrids associated with an adjacent anion-cation pair. V_{1A} (V_{1C}) is the interaction between two hybrids on the same anion (cation), the "constant energy" V_2 is the interaction between anion and cation hybrids that point toward one another along the bond direction, V_1^x is the interaction between an anion hybrid in one direction and an adjacent cation hybrid pointing in a different direction. The "ionic energy" V_3 is half the difference between cation and anion term values. The lower segment of the figure depicts the splitting of the hybrid energy levels by the V_2 , and V_3 interactions.
3. A schematic picture of the hybrids, treated as rigid, in a shear distortion leading to $C_{11} - C_{12}$.
4. List of possible crystal structures for ordered semiconductor alloys (Wei and Zunger, 1989).
5. Near neighbor bond lengths (GaAs and InAs in the $Ga_{1-x}In_xAs$ alloy as a function of composition x , measured by EXAFS (Mikkelsen and Boyce, 1982).
6. Cluster populations relative to those in a random alloy $x_j - \bar{x}_j$ for clusters with $n_j = 0, 1, 2, 3, 4$ B atoms for a $Ga_{1-x}In_xAs$ alloy equilibrated at 600 K.
7. Atom configurations during the slip of a dislocation.

8. Schematic of the dislocation potential as a function of its position.
9. Measured hardness of Ge and several GaAs samples as a function of temperature.

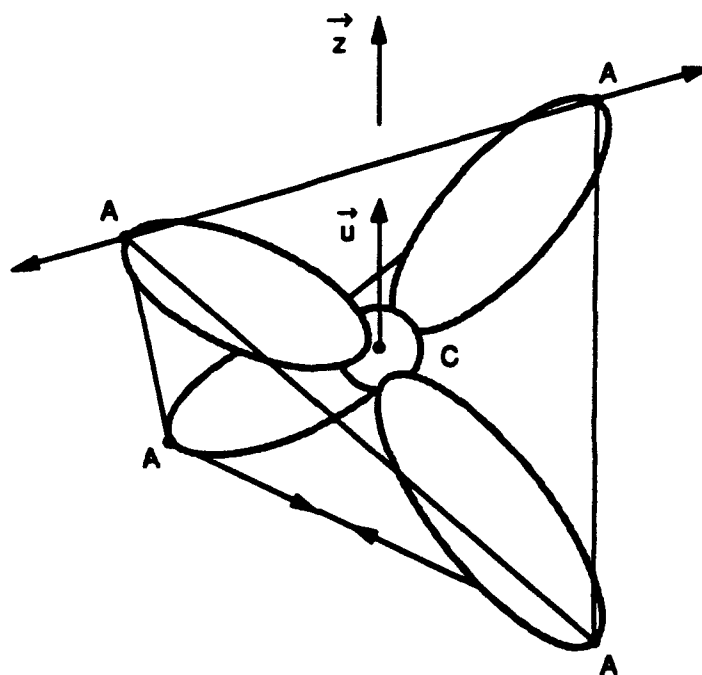
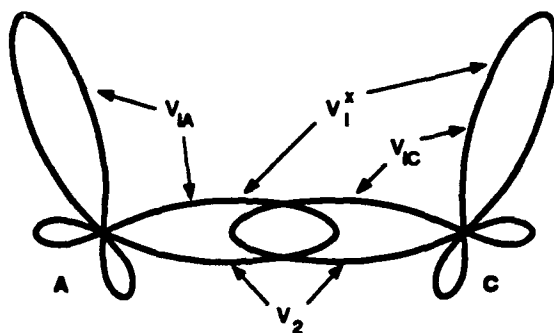
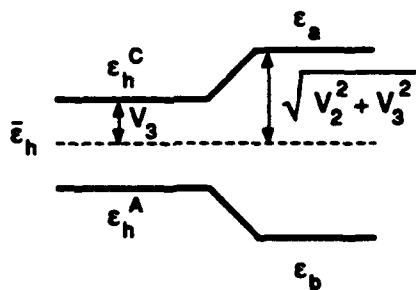


Figure 1



$$V_3 = 1/2 (\epsilon_h^C - \epsilon_h^A)$$



mh-090790-jd

Figure 2

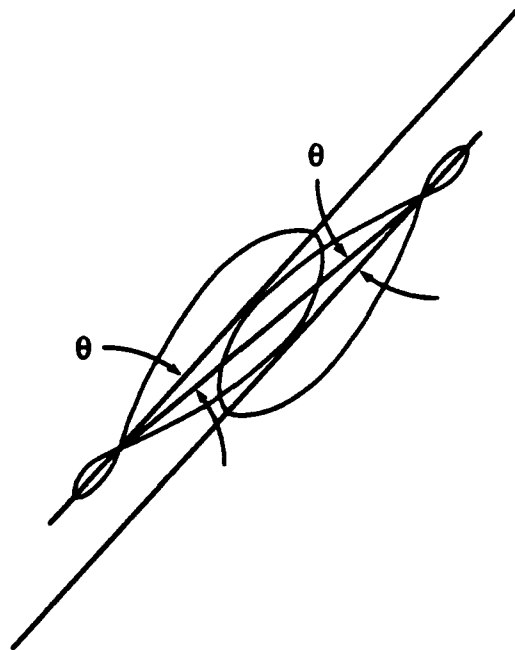
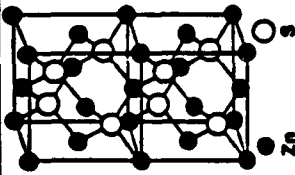
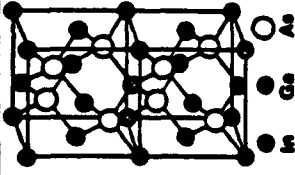
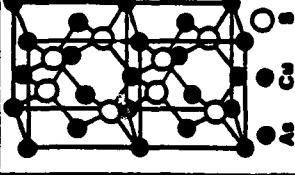
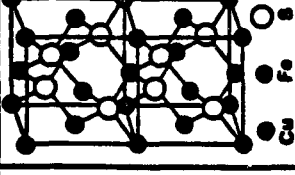
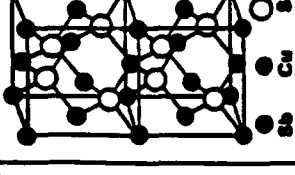
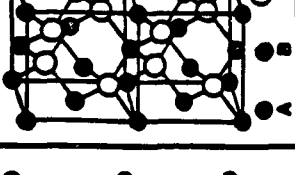
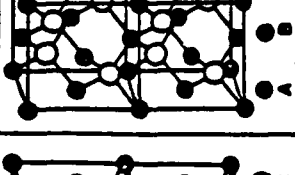
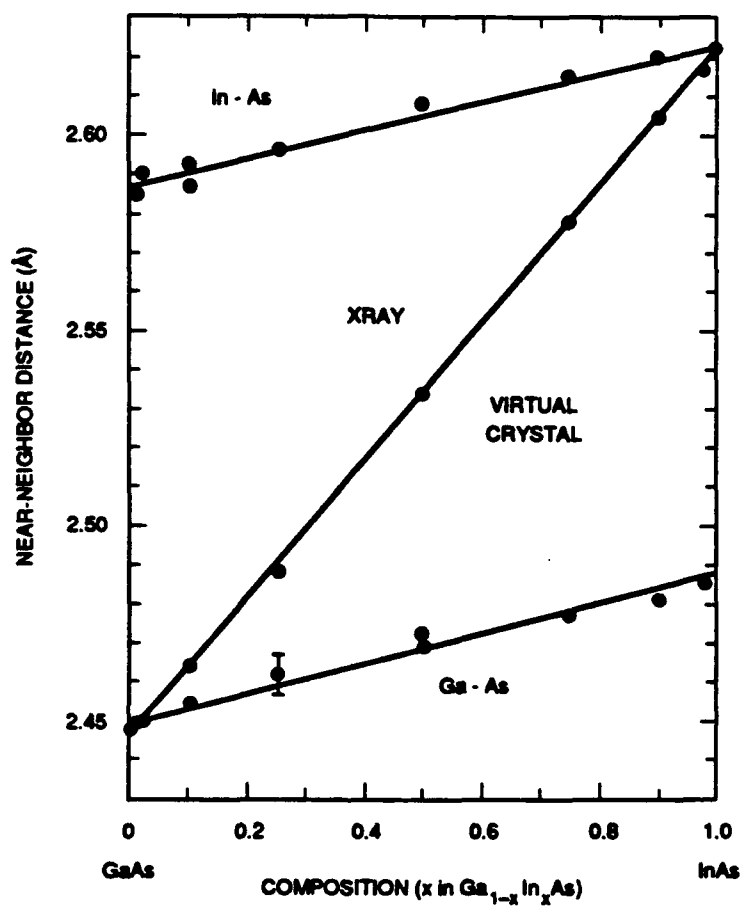


Figure 3

Ordering Vectors		(0,0,0)		(0,0,1)		(2,0,1)		(1,1,1)	
Name (ternary)		Zincblende (sphalerite)		Layered Tetragonal	"Luzonite"	Chalcopyrite	Ferriantite	Layered Trigonal	"X"
Formula		$n = 0,4,AC$		$n = 2,ABC_2$	$n = 1,2,ABC_2$	$n = 2,ABC_2$	$n = 1,2,ABC_2$	$n = 2,ABC_2$	$n = 1,2,ABC_2$
									
		Zn S		In Ga As	As Cu S	Cu Fe S	Sb Cu S	A B C	A B C
Examples (ternary)		ZnS-type		InGaAs ₂ -type	Cu ₃ AsS ₄ -type	CuFeS ₂ -type	Cu ₃ SbS ₄ -type	CrCuS ₂ -type (Na V S ₂)	
Bravais Lattice		Face centered cubic		Simple tetragonal	Simple cubic	Body centered tetragonal	Body centered tetragonal	Rhombohedral	Orthorhombic
Space Group		$Fm\bar{3}m$		$P4_2/m$	$P\bar{2}3m$	$P4_2/m$	$P4_2/m$	$R\bar{3}m$	
Int. Tables		T_d^2		D_{2d}^2	T_d^1	D_{2d}^2	D_{2d}^2	C_{3v}	
Schönflies		T_d		D_{2d}	T_d	D_{2d}	D_{2d}	C_{3v}	
Number		216		118	216	122	121	100	
Strukturbericht		B3		—	H2 ₄	E1 ₁	H2 ₄	—	
Pearson symbol		CF8		—	CP8	816	816	816	
Atomic positions (ternary)		Zn: 4 a 2m S: 4 c 2m		1A: 1 a 2m 1B: 1 c 2m 2C: 2 b mm	3Cu: 3 a 2m 1As: 1 a 2m 4S: 4 c 3m	2Cu: 4 a 2 2Fe: 4 b 2 4S: 8 d 2	1Cu: 2 b 2m 2Cu: 4 d 2 1Sb: 2 c 2m 4S: 8 l m	1A: 1 a 3m 1B: 1 c 3m 1C: 1 a 3m 1C: 1 c 3m	
Equivalent superlattices		None		(1,1) in (001) direction	None	(2,2) in (201) direction	(1,2) in (201) direction	(1,1) in (111) direction	None
Space group and type (binary)		$Fm\bar{3}m$		$P4_2/m$	$P\bar{2}3m$	$P4_2/m$	$P4_2/m$	$R\bar{3}m$	Common
Int. Tables		O_h^8		D_{2d}^2	O_h^1	D_{2d}^2	D_{2d}^2	D_{3d}^5	10
Schönflies		O_h		D_{2d}	O_h	D_{2d}	D_{2d}	D_{3d}	24
Number		228		122	221	141	130	100	88
Strukturbericht		A1		L1 ₂	L1 ₂	—	DO 20	L1 ₁	—
Pearson symbol		CF ₄		164	CP4	88	88	832	—
Examples (binary)		Cu ₃ Ag ₃ As		Cu ₃ As	Cu ₃ As As ₃ Cu	None	Al ₃ Tl Ni ₃ V	Cu ₃ Pt SiO ₂	None



sh-090690-jd

Figure 5

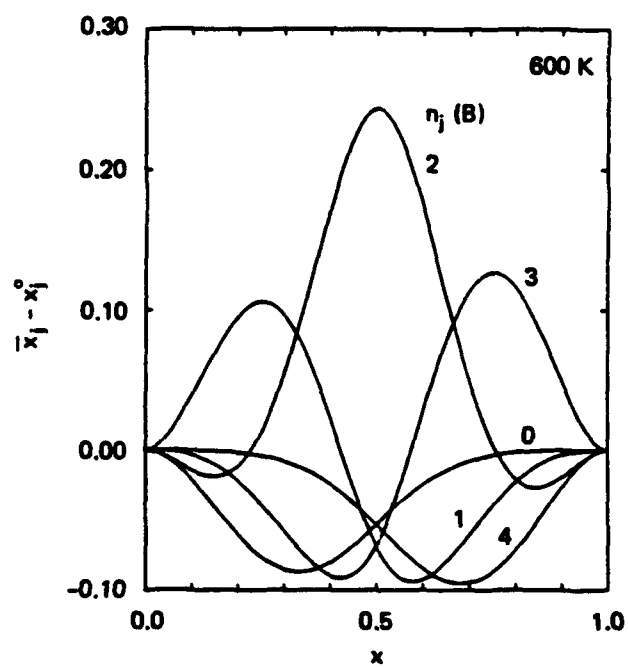
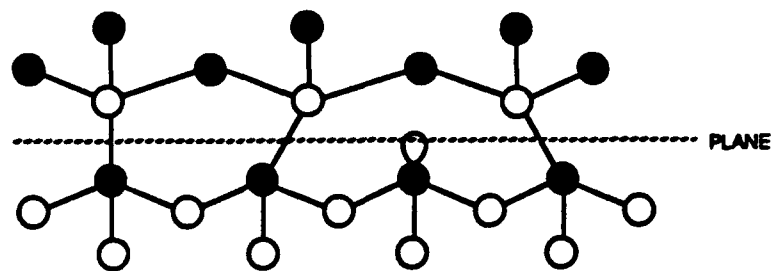
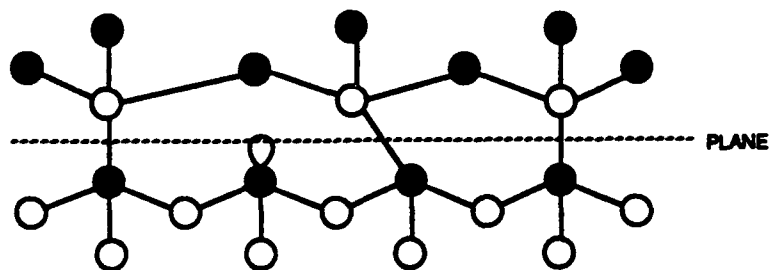


Figure 6



(a) CONFIGURATION A



(b) CONFIGURATION B

bo-001090-jd

Figure 7

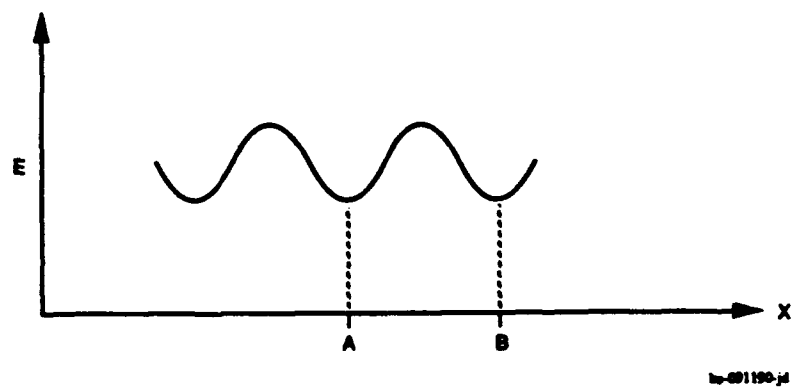


Figure 8

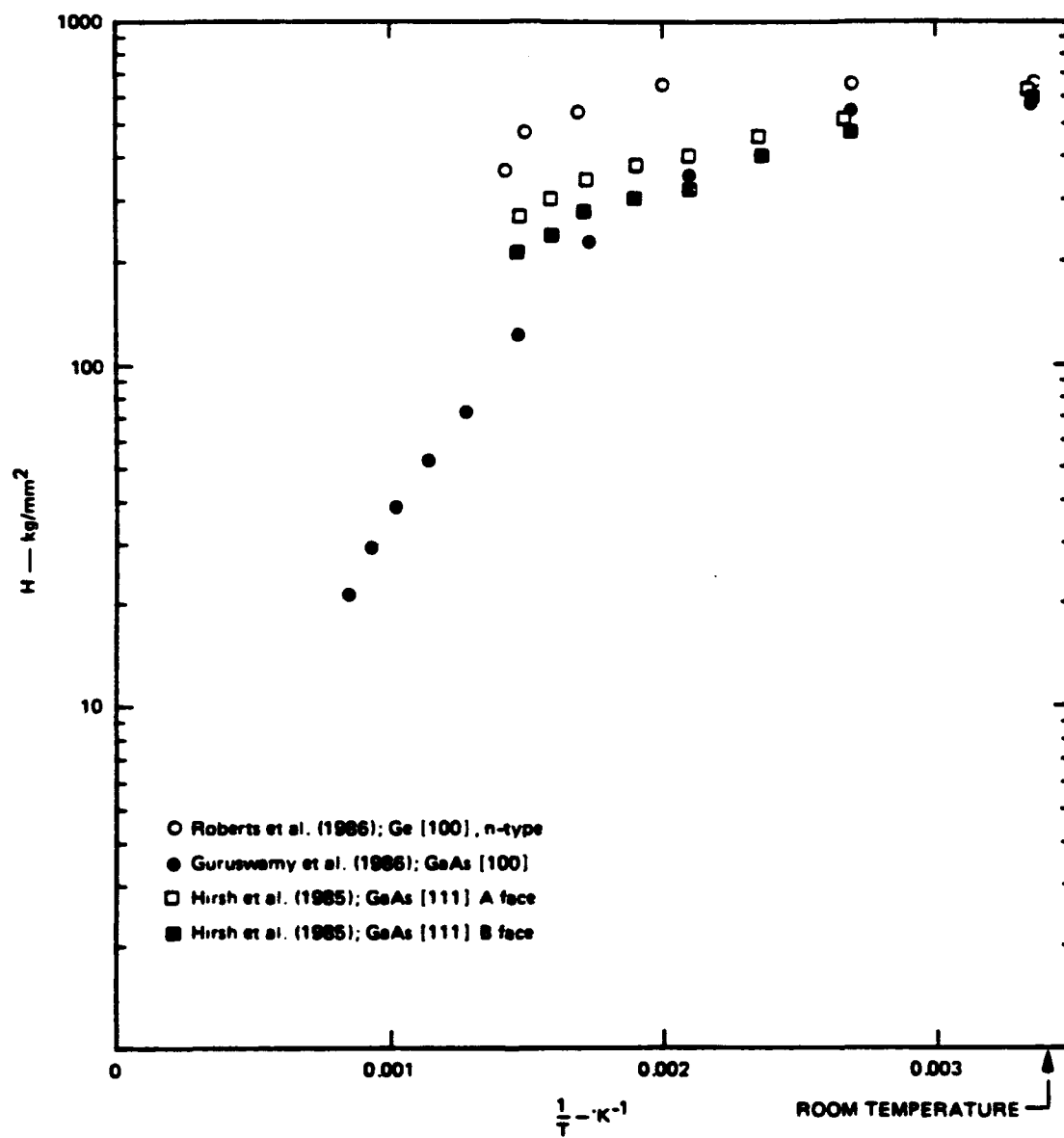


Figure 9

Table 1. Comparison between calculated experimental constants.^a

	Expt	FP-LMTO ^b	PP-PW ^c	TB
Si				
a	5.431	5.41	5.45	5.431
B	9.923	9.9	9.3	9.923
C ₁₁ - C ₁₂	10.274	10.2	9.8	10.274
C ₄₄	8.036	8.3	8.5	8.013
C ₄₄ ^(o)			11.1	11.30
ζ	0.54 ^d	0.51	0.53	0.51
ω	523	518	521	572
Ge				
a	5.65		5.59	5.65
B	7.653		7.2	7.653
C ₁₁ - C ₁₂	8.189		8.5	8.189
C ₄₄	6.816		6.3	6.84
C ₄₄ ^(o)			7.7	9.46
ζ			0.44	0.49
ω	303		302	342

Table 1. (Continued)

GaAs

a	5.642	5.55	5.642
B	7.69	7.3	7.69
C ₁₁ - C ₁₂	6.63	7.0	6.63
C ₄₄	6.04	6.2	5.79
C ₄₄ ^(o)		7.5	7.83
ζ		0.48	0.50
ω	273	268	292

a. Comparison between calculated and experimental lattice constant a , elastic constants B , $C_{11} - C_{12}$ and C_{44} , Kleiman (1962) internal distortion parameter ζ , and the TO optical phonon ω in wave numbers 1/cm. Also listed are $C_{44}^{(o)}$ defined in Eq. (13). The FP-LMTO and PP-PW are the *ab initio* theories described in Part I and TB is the tight-bonding theory in Part III, Section 10. All elastic constants are in units of 10^{11} dynes/cm².

b. Methfessel et al. (1989)

c. Nielsen and Martin (1985a)

d. Cousins et al. (1987)

e. All the experimental lattice constants are those tabulated by Zallen (1982). The experimental elastic constants are taken from Table 3, and the phonon frequencies are from Table 5.

Table 2. Accuracies of measurement techniques.

Method	Features	Accuracy	Corrections
Pulsed			
Pulse Echo	<ul style="list-style-type: none"> • cm or longer samples • one transducer • commercial units available 	a few parts in 1000	none
Gated Double Pulse Superposition ^a	<ul style="list-style-type: none"> • cm or longer samples • matched transducers 	1 part in 10,000	bond
Pulse Superposition ^{b,c}	<ul style="list-style-type: none"> • cm or longer samples • one transducer 	1 part in 5000 ^a	bond, diffraction estimate (b)
Pulse Echo Overlap ^{d,e}	<ul style="list-style-type: none"> • cm or longer samples • one transducer • commercial units available 	(delay time accuracy several parts per 1,000,000)	bond, diffraction
Long Pulse Buffer Rod ^f	<ul style="list-style-type: none"> • thin lossy samples (as thin as 2 mm) 	1 or 2 parts in 10,000	bond
Continuous Wave	<ul style="list-style-type: none"> • bonds must be thin • cm or longer length • sample preparation more critical 		requires corrections for all results. Otherwise, errors can be very large. ^p Difficult to correct for diffraction. All these techniques require compound resonator corrections.
Continuous Wave Transmission ^{g,h,i}	<ul style="list-style-type: none"> • can be automated • electromagnetic cross talk can give problems 	parts in 10,000,000 under ideal conditions; considerably less with thick bonds and/or thin samples ^k	

Table 2. (Continued)

Sampled Continuous Wave Method ^{l,k}	<ul style="list-style-type: none"> • Single transducer • eliminates electromagnetic crosstalk • can be automated 		
Spectrometer Methods ^{l,m}	<ul style="list-style-type: none"> • can track small changes • can require specialized transducers 		
Optical Methods (for use with optically transparent media)			
Measurement of wavelength	<ul style="list-style-type: none"> • uses travelling microscope to measure wavelength 	0.1 to 0.01%	
Raman-Nath Diffraction	<ul style="list-style-type: none"> • measurement of angles of incidence and diffraction to measure wavelength of sound 	1%	
Bragg Diffraction	<ul style="list-style-type: none"> • useful at frequencies between ~100 MHz and several GHz • useful for measurement of local variations in velocity 	0.01 to 0.1%	
Brillouin Scattering	<ul style="list-style-type: none"> • measurement of optical frequency shift and angle • useful near phase transitions • frequencies in the GHz range 	1. to 3% (0.1% precision)	
Stimulated Brillouin Scattering ^{n,o}	<ul style="list-style-type: none"> • needs high power electromagnetic source • measurement can result in fracture of sample • frequencies to 60 GHz reported 	3%	

- | | | | |
|-----------------------------|----------------------------|--------------------------|--------------------------------|
| a. Williams and Lamb (1985) | e. Papadakis (1972) | i. Chern et al. (1981) | m. Heyman (1976) |
| b. McSkimin (1961) | f. McSkimin (1950) | j. Miller (1973) | n. Chiao et al. (1964) |
| c. Papadakis (1972) | g. Bolef et al. (1962) | k. Chern et al. (1981) | o. Brewer and Fleckhoff (1964) |
| d. Papadakis (1967) | h. Bolef and Miller (1974) | l. Conradi et al. (1974) | p. Chern et al. (1981) |

Table 3. Experimental values for cubic semiconductors.^a

	C ₁₁	C ₁₂	C ₄₄	s	α	β	l_k	l_M	l_{BOM}
C ^b	107.640	12.520	57.740	0.0	129.100	84.573	1.00	1.00	1.02
Si ^b	16.772	6.498	8.036	0.0	49.247	13.951	1.00	1.00	1.13
Ge ^b	13.112	4.923	6.816	0.0	39.438	11.583	1.08	1.08	1.05
AlSb ^b	8.769	4.341	4.076	1.684	33.768	6.653	1.11	1.05	1.08
GaP ^b	14.390	6.520	7.143	3.815	46.965	10.448	1.12	1.05	1.08
GaAs ^b	12.110	5.480	6.040	2.827	40.895	9.159	1.12	1.06	1.05
GaSb ^b	9.089	4.143	4.440	1.569	33.123	7.412	1.10	1.06	1.06
InP ^b	10.220	5.760	4.600	3.766	41.095	6.250	1.20	1.07	1.03
InAs ^b	8.329	4.526	3.959	2.820	33.744	5.531	1.22	1.11	1.00
InSb ^b	6.918	3.788	3.132	1.372	29.909	4.951	1.17	1.11	1.05
ZnS ^c	9.420	5.680	4.360	6.788	37.026	4.571	1.33	1.07	0.95
ZnS ^c	10.790	7.220	4.120	6.788	45.126	4.341	1.28	1.02	1.01
ZnS ^c	9.810	6.270	4.483	6.788	39.947	4.300	1.42	1.13	0.90
ZnS ^c	10.460	6.530	4.630	6.788	41.880	4.828	1.33	1.08	0.95
ZnSe ^b	8.95	5.39	3.984	4.368	34.432	4.716	1.28	1.09	0.98
ZnSe ^c	8.59	5.06	4.06	4.368	34.519	4.673	1.32	1.13	0.95
ZnSe ^c	8.720	5.240	3.920	4.368	35.469	4.603	1.29	1.10	0.98
ZnTe ^c	7.130	4.070	3.120	2.566	29.976	4.452	1.18	1.06	1.05
ZnTe ^c	7.220	4.090	3.080	2.566	30.204	4.558	1.14	1.03	1.08
CdTe ^b	5.33	3.65	2.04	3.105	27.058	2.455	1.34	1.07	0.98

Table 3. (Continued)

CdTe ^c	6.150	4.300	1.960	3.105	31.546	2.731	1.16	0.94	1.13
HgTe ^b	5.971	4.154	2.259	2.381	30.300	2.542	1.37	1.16	0.96
HgTe ^c	5.63	3.66	2.11	2.381	26.919	2.542	1.37	1.15	0.95

a. Experimental elastic constants for some cubic semiconductors and the parameter s of Eq. (81) taken from Martin (1970) with the force constants α and β obtained from Eqs. (82) and (83) and the identity relations l_K , l_M and l_{BOM} given by Eqs. (78), (88), and (119), respectively. The elastic constants C are in units of 10^{11} dynes/cm², and the remainder are in 10^3 dynes/cm.

b. Data quoted from "Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology," New Series, Vols. 17 and 22.

c. Listed in the review by Mitra and Massa (1982).

Table 4. Values for the term values used in the present calculations and Mann's (1967) Hartree-Fock values as used by Harrison (1980).

Element	Present		Mann	
	ϵ_s	ϵ_p	ϵ_s	ϵ_p
Cu	-7.72	-2.96	-7.72	-2.37
Ag	-7.57	-3.54	-7.06	-2.61
Au	-9.22	-3.91	-6.98	-2.67
Be	-9.32	-5.41	-8.41	-5.79
Mg	-7.62	-2.97	-6.88	-3.84
Zn	-9.39	-4.09	-7.96	-4.02
Cd	-8.99	-4.17	-7.21	-3.99
Hg	-10.43	-4.35	-7.10	-3.95
B	-14.00	-8.30	-13.46	-8.43
Al	-11.78	-5.98	-10.70	-5.71
Ga	-13.23	-5.90	-11.55	-5.67
In	-12.03	-5.56	-10.14	-5.37
C	-19.81	-11.26	-19.37	-11.07
Si	-15.03	-8.15	-14.79	-7.58
Ge	-16.40	-7.75	-15.15	-7.33
Sn	-14.53	-7.03	-13.04	-6.76
Pb	-15.25	-6.45	-12.48	-6.53
N	-26.08	-14.54	-26.22	-13.84
P	-19.62	-10.57	-19.22	-9.54
As	-20.02	-9.93	-18.91	-8.98
Sb	-17.56	-8.77	-16.02	-8.14

Table 4. (Continued)

O	-28.55	-13.561	-34.02	-16.72
S	-21.16	-10.39	-24.01	-11.60
Se	-21.41	-9.90	-22.86	-10.68
Te	-19.12	-9.32	-19.12	-9.54
F	-36.23	-17.44	-42.78	-19.86
Cl	-25.81	-13.05	-29.19	-13.78
Br	-24.95	-12.01	-27.00	-12.43
I	-21.95	-10.79	-22.34	-10.97

Table 5. Experimental Properties of Semiconductors.^a

	d	E _{bond}	B	C	C ₄₄	ω
C	1.540	-3.68	44.227	95.120	57.740	1332
Si	2.352	-2.32	9.923	10.274	8.036	520
Ge	2.450	-1.94	7.653	8.189	6.816	301
AlP	2.367	-2.13	8.600	6.900	6.150	440
AlAs	2.451	-1.89	7.727	7.160	5.420	361
AlSb	2.656	-1.76	5.817	4.428	4.076	366
GaP	2.360	-1.78	9.143	7.780	7.143	367
GaAs	2.448	-1.63	7.690	6.630	6.040	269
GaSb	2.640	-1.48	5.792	4.946	4.440	231
InP	2.541	-1.74	7.247	4.460	4.600	304
InAs	2.622	-1.55	5.794	3.803	3.959	219
InSb	2.805	-1.40	4.831	3.130	3.132	185
Zn S	2.342	-1.59	7.637	3.990	4.558	279
ZnSe	2.454	-1.29	6.457	3.560	3.984	213
ZnTe	2.637	-1.20	5.090	3.060	3.120	177
CdTe	2.806	-1.10	4.210	1.680	2.040	141
HgTe	2.798	-0.81	4.759	1.817	2.259	116

a. Values of bond length d, bond energy E_{bond}, bulk modulus B and shear coefficient C = C₁₁-C₁₂ used to determine the parameters in Tables 7 through 11. Also listed are the experimental values of C₄₄ and the TO optical phonon mode ω at Γ to be compared with the calculations. All the elastic constants are in units of 10¹¹ dynes/cm², d in Å, E_{bond} in eV, and ω_{TO} in terms of wave numbers in 1/cm.

Table 5. (Continued)

All bond lengths are deduced from the lattice constants quoted by Zallen (1982). The values of E_{bond} are taken from Harrison (1980), Table 7-3, except for AlSb, ZnTe, CdTe, and HgTe, which are deduced from the Phillips (1973) Table 8.2. The elastic constants are taken from Table 3, and the phonon frequencies are taken from values compiled in "Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology," New Series, edited by K.-H. Hellwedge, Vols. 17 and 22.

Table 6. Comparison of BOM and BS Models' Predictions.^a

	<u>E_{bonds}</u>		<u>B</u>		<u>C</u>		<u>C₄₄</u>	
	BOM	BS	BOM	BS	BOM	BS	BOM	BS
C	-7.17	-7.08	49.31	47.39	67.15	67.10	44.45	51.78
Si	-2.39	-2.50	4.65	4.18	7.19	8.07	5.10	5.29
Ge	-2.03	-2.34	2.89	1.91	4.72	6.58	3.69	3.00
AlP	-2.59	-2.54	3.81	4.21	4.45	4.83	3.45	4.10
AlAs	-2.22	-2.20	3.12	3.66	3.85	4.07	2.91	3.50
AlSb	-1.64	-1.62	1.99	2.47	2.75	3.12	2.11	2.54
GaP	-2.32	-2.16	3.63	4.40	4.48	5.15	3.69	4.34
GaAs	-1.99	-1.85	2.83	3.76	3.75	4.30	3.02	3.66
GaSb	-1.53	-1.43	1.76	2.45	2.63	3.40	2.17	2.65
InP	-2.21	-2.06	2.35	2.91	2.69	2.93	2.21	2.63
InAs	-1.89	-1.74	1.93	2.67	2.38	2.52	1.90	2.32
InSb	-1.43	-1.28	1.28	1.97	1.81	2.17	1.48	1.87
ZnS	-1.83	-1.79	3.46	3.74	3.53	3.62	2.71	3.30
ZnSe	-1.52	-1.48	2.67	3.06	2.75	2.67	2.08	2.53
ZnTe	-1.15	-1.05	1.82	2.16	1.88	1.87	1.50	1.78
CdTe	-1.06	-0.97	1.24	1.47	1.22	1.14	0.96	1.13
HgTe	-0.72	-0.49	1.23	1.72	1.33	1.31	1.11	1.30

a. Comparison of the tight-binding theory using the full band structures (BS) and the bond orbital model (BOM) for bond energies E_{bond} , bulk moduli B , and shear coefficients $C = C_{11} - C_{12}$ and C_{44} . All energies are in eV and elastic constants in 10^{11} dynes/cm².

Table 7. Full band structure calculation.^a

	f	n	m	u_0	C_{44}	ζ	ω
C	1.390	2.840	3.767	21.924	48.393	0.121	1459
Si	1.326	3.040	5.001	6.938	8.013	0.511	572
Ge	1.388	3.204	5.278	6.415	6.841	0.487	342
AlP	1.294	3.530	5.598	6.435	5.827	0.516	447
AlAs	1.464	3.524	5.430	7.089	5.598	0.459	384
AlSb	1.337	3.268	5.668	4.838	3.944	0.564	354
GaP	1.395	3.705	5.683	7.285	6.857	0.501	382
GaAs	1.397	3.633	5.716	6.530	5.791	0.500	292
GaSb	1.431	3.471	5.717	5.519	4.515	0.536	256
InP	1.323	4.240	6.633	5.603	4.260	0.584	304
InAs	1.300	3.997	6.427	4.962	3.564	0.552	220
InSb	1.353	3.773	6.399	4.350	3.092	0.602	200
Zn S	1.062	3.308	5.996	4.225	3.727	0.632	325
ZnSe	1.134	3.420	5.994	4.260	3.164	0.576	233
ZnTe	1.284	3.3096	5.828	4.285	2.813	0.590	205
CdTe	1.171	3.656	6.761	3.092	1.701	0.694	156
HgTe	1.173	3.760	7.074	3.080	2.040	0.716	152

a. The results for the parameters f, n, m, and u_0 obtained from the fitting of the bond energy, bond length, bulk modulus, and shear coefficient $C_{11} - C_{12}$ of Table 5 using the full band structure calculation. Also listed are the calculated C_{44} , internal displacement parameter ζ , and the TO optical phonon mode ω at Γ . All the elastic constants are in units of 10^{11} dynes/cm², u_0 is in eV, and ω are wave numbers in 1/cm.

Table 8. Two sets of TB parameters.^a

Si									
	ϵ_s^A	ϵ_p^A	ϵ_s^C	ϵ_p^C	$V_{ss\sigma}$	$V_{sp\sigma}$	$V_{pp\sigma}$	$V_{pp\pi}$	
Chadi	0.0	7.20	0.0	7.20	-2.03	2.55	4.55	-1.09	
Present	0.0	6.88	0.0	6.88	-2.41	2.59	4.05	-1.15	
	u_0	u_1	u_2	C	C_{44}	$C_{44}^{(o)}$	ζ	ω	
Chadi	7.29	-9.98	23.90	10.66	7.89	11.38	0.49	620	
Present	6.93	-9.70	23.42	10.27	7.83	11.39	0.51	592	
GaAs									
	ϵ_s^A	ϵ_p^A	ϵ_s^C	ϵ_p^C	$V_{ss\sigma}$	$V_{sp\sigma}^{AC}$	$V_{sp\sigma}^{CA}$	$V_{pp\sigma}$	$V_{pp\pi}$
Chadi	0.0	9.64	5.12	11.56	-1.70	2.40	1.90	3.44	-0.89
Present	0.0	10.09	6.79	14.12	-2.34	2.52	2.52	3.94	-1.12
	u_0	u_1	u_2	C	C_{44}	$C_{44}^{(o)}$	ζ	ω	
Chadi	5.12	-7.12	18.22	6.36	5.60	8.77	0.54	339	
Present	6.53	-8.39	19.90	6.63	5.70	8.53	0.54	322	

a. Comparison between the two different sets of TB parameters described in the text, the resultant expression coefficients u_0 , u_1 , and u_2 of the repulsive pair energy u , and the predicted elastic constants, Kleinmann internal displacement parameters ζ , and phonon frequency ω from Chadi fitting scheme.

Table 9. BOM(1) calculations.^a

	f	n	m	u_0	C_{44}	ζ	ω
C	1.440	2.896	3.809	22.345	61.098	0.142	2103
Si	1.356	3.208	5.166	7.023	9.729	0.447	695
Ge	1.395	3.666	5.854	6.383	8.240	0.506	416
AlP	1.047	3.334	5.842	4.418	6.627	0.460	571
AlAs	1.264	3.530	5.685	5.538	6.668	0.431	491
AlSb	1.179	3.170	5.831	3.818	4.482	0.511	411
GaP	1.154	3.439	5.744	5.239	7.500	0.453	496
GaAs	1.179	3.337	5.647	4.863	6.340	0.456	358
GaSb	1.274	3.354	5.797	4.456	4.950	0.500	302
InP	0.999	3.579	6.490	3.413	4.558	0.522	388
InAs	0.995	3.167	5.960	3.081	3.787	0.496	260
InSb	1.126	3.228	6.105	3.091	3.267	0.543	224
Zn S	0.692	2.750	7.262	1.642	4.312	0.580	401
ZnSe	0.750	2.823	6.839	1.821	3.685	0.544	287
ZnTe	0.888	2.715	6.484	1.987	3.156	0.531	248
CdTe	0.734	2.720	8.394	1.018	1.949	0.660	189
HgTe	0.732	2.202	8.810	0.842	2.140	0.677	177

a. The results for the parameters f , n , m , and u_0 obtained from the fitting of the bond energy, bond length, bulk modulus and shear coefficient C_{11} - C_{12} of Table 5 using the BOM(1) described in the text. Also listed are the calculated C_{44} , internal displacement parameter ζ , and the TO optical phonon mode ω at Γ . All the elastic constants are in units of 10^{11} dynes/cm², u_0 is in eV, and ω is given in terms of wave number in 1/cm.

Table 10. BOM(2) calculations.^a

	f	n	m	u_0	C_{44}	ζ	ω_{TO}
C	1.440	2.896	3.809	22.345	55.161	0.319	1672
Si	1.356	3.208	5.166	7.023	7.520	0.652	597
Ge	1.395	3.666	5.854	6.383	6.106	0.720	358
AlP	1.283	3.440	5.493	6.171	5.094	0.677	511
AlAs	1.472	3.543	5.419	6.979	5.103	0.640	440
AlSb	1.342	3.249	5.580	4.771	3.256	0.706	373
GaP	1.371	3.540	5.490	6.901	5.916	0.688	428
GaAs	1.387	3.455	5.452	6.306	4.864	0.681	317
GaSb	1.414	3.398	5.593	5.305	3.681	0.717	264
InP	1.307	3.846	6.115	5.339	3.446	0.744	354
InAs	1.298	3.528	5.731	4.834	2.806	0.706	243
InSb	1.342	3.464	5.910	4.204	2.351	0.747	207
Zn S	1.062	3.102	5.728	4.075	3.016	0.726	373
ZnSe	1.141	3.150	5.587	4.181	2.577	0.688	271
ZnTe	1.283	3.067	5.498	4.140	2.270	0.704	231
CdTe	1.178	3.271	6.216	3.011	1.319	0.764	177
HgTe	1.169	3.046	6.116	2.916	1.457	0.793	162

a. The results for the parameters f , n , m , and u_0 obtained from the fitting of the bond energy, bond length, bulk modulus and shear coefficient C_{11} - C_{12} of Table 5 using the BOM(2) described in the text. Also listed are the calculated C_{44} , internal displacement parameter ζ , and the TO optical phonon mode ω at Γ . All the elastic constants are in units of 10^{11} dynes/cm², u_0 is in eV, and ω is wave number in 1/cm.

Table 11. BOM(3) calculations.^a

	f	n	m	u ₀	C ₄₄	ζ	ω
C	1.336	2.901	3.872	20.814	47.691	0.135	1531
Si	1.262	3.251	5.346	6.484	7.472	0.580	562
Ge	1.302	3.886	6.264	5.942	6.171	0.658	333
AlP	1.193	3.43	5.577	5.722	4.936	0.541	452
AlAs	1.369	3.559	5.507	6.498	4.812	0.490	387
AlSb	1.249	3.253	5.669	4.402	3.248	0.619	343
GaP	1.276	3.550	5.575	6.422	5.738	0.542	373
GaAs	1.290	3.445	5.511	5.855	4.724	0.549	281
GaSb	1.318	3.415	5.704	4.923	3.670	0.621	239
InP	1.210	3.812	6.153	4.929	3.398	0.604	304
InAs	1.199	3.456	5.724	4.442	2.755	0.579	216
InSb	1.245	3.411	5.943	3.871	2.367	0.657	188
Zn S	0.968	3.062	5.848	3.618	3.120	0.630	343
ZnSe	1.040	3.097	5.671	3.727	2.611	0.580	247
ZnTe	1.178	3.019	5.573	3.371	2.301	0.601	212
CdTe	1.065	3.198	6.368	2.621	1.403	0.690	165
HgTe	1.059	2.937	6.262	2.523	1.563	0.730	154

a. The results for the parameters f, n, m, and u₀ obtained from the fitting of the bond energy, bond length, bulk modulus and shear coefficient C₁₁-C₁₂ of Table 5 using the BOM(3) described in the text. Also listed are the calculated C₄₄, internal displacement parameter ζ, and the TO optical phonon mode ω at Γ. All the elastic constants are in units of 10¹¹ dynes/cm², u₀ is in eV, and ω is wave number in 1/cm.

Table 12. Ordered alloys: TB elastic constant calculations.^a

Alloy	B_{Ch}	B_{Ca}	B_{Cp}	\bar{B}	$\Delta B/\bar{B} \times 100$		
					Ch	Ca	Cp
AlGaAs	7.695	7.693	7.689	7.909	-0.18	-0.20	-0.25
AlGaP	8.858	8.858	8.854	8.872	-0.15	-0.15	-0.20
GaInSb	5.226	5.202	5.156	5.312	-1.61	-2.07	-2.92
AlInAs	6.705	6.691	6.661	6.761	-0.83	-1.03	-1.48
InGaAs	6.610	6.579	6.508	6.742	-1.96	-2.42	-3.47
InAlP	7.876	7.860	7.774	7.924	-0.61	-0.08	-1.88
GaInP	8.007	8.035	8.878	8.195	-1.44	-1.95	-3.87
GaAsP	8.328	8.291	8.294	8.417	-1.05	-1.50	-1.46
GaAsSb	6.314	6.198	6.157	6.741	-6.34	-8.05	-8.66
GaPSb	6.584	6.297	6.188	7.468	-11.84	-15.68	-17.14
HgCdTe	4.470	4.472	4.471	4.485	-0.33	-0.28	-0.31
HgZnTe	4.890	4.887	4.632	4.925	-0.71	-0.76	-5.93
CdZnTe	4.611	4.604	4.338	4.650	-0.85	-1.00	-6.72

a. Bulk moduli of ordered alloys calculated using the full TB band-structure method described in Part III, Section 10. The three structures are chachopyrite (Ch), CuAul (Ca) and CuPt(Cp) types. \bar{B} is the average value of the constituent compounds, and $\Delta B = B - \bar{B}$.

Table 13. Ordered alloys: Bulk moduli.^a

	<u>GaAs</u>	<u>GaSb</u>	<u>Ga₂AsSb</u>			<u>Ga₄As₃Sb</u>		<u>Ga₄AsSb₃</u>	
1 ₂ structure	zb	zb	CA	CH	CP	LU	FA	LU	FA
B	7.46	5.18	6.10	5.92	5.96	6.52	6.58	5.40	5.31
$\Delta B/\bar{B} \times 100$			-3.5	-6.3	-5.7	-5.4	-4.5	-6.1	-7.7

a. Calculated bulk moduli for GaAs, GaSb, and GaAsSb ordered alloys by Ferreira et al. (1989), and the corresponding percentage deviation from the concentration weighed average. The structures are zincblende (zb), CuAl(CA), chachopyrite (CH), CuPt(CP), Luzonite (LU), and Farnatinite (FA).

Table 14. Bulk moduli of ordered binary alloys AB of the diamond semiconductors A and B from theories and experiment.

SIC	PP-PW ^a	experiment		
B(C)	50.3	44.23		
B(Si)	9.53	9.92		
B(SiC)	23.4	22.4		
$\Delta B/\bar{B}(\%)$	-21	-17		
SIG ^e	PP-PW ^a	PP-PW ^b	ASAC ^c	FP-LMTO ^c
B(Si)	9.53	9.8	8.80	9.58
B(Ge)	7.75	7.7	6.25	7.05
B(SiGe)	8.73	8.7	7.38	8.31
$\Delta B/\bar{B}(\%)$	1	0	2	0

a. Martins and Zunger (1986)

b. Oteish and Resta (1988)

c. van Schilfgaarde (1990)

Table 15. Measured elastic constants in 10^{11} dynes/cm² of SiGe alloys by Bublik et al. (1974)..

alloy	C ₁₁	C ₁₂	C ₄₄
Si _{0.28} Ge _{0.72}	16.1 ± 0.8	8.35 ± 0.8	8.55 ± 0.4
Si _{0.54} Ge _{0.46}	17.0 ± 0.8		
Si _{0.64} Ge _{0.36}	17.1 ± 0.8		

Table 16. Calculated Peierls stress and hardness for various zincblende semiconductors, with experimental hardness values for comparison.^a

	U_b	τ_p	H_p	H_{Int}	$H_p + H_{Int}$	H_{exp}
C	11.35	23,300	1940	9244	11,184	1
Si	5.95	3430	286	1098	1384	1370
Ge	6.67	3440	286	893	1179	1000
Sn	5.66	1930	161	—	—	—
AlP	6.04	3410	284	—	—	—
GaP	6.15	3500	292	903	1195	940
InP	5.78	2640	220	548	768	520
AlAs	5.90	3000	250	—	—	505
GaAs	6.03	3000	256	750	1006	580
InAs	5.64	2340	195	469	664	430
AlSb	5.08	2840	237	524	761	400
GaSb	5.36	2180	182	553	735	450
InSb	4.92	1670	139	365	504	230
ZnS	5.14	3000	250	515	765	—
CdS	4.77	2750	229	288	517	—
HgS	5.68	2590	216	—	—	—
ZnSe	4.92	2500	209	462	671	137
CdSe	4.59	1890	157	254	411	—

Table 16. (Continued)

HgSe	4.523	1840	154	232	386	—
ZnTe	4.42	1720	143	374	517	82
CdTe	4.01	1350	113	222	335	60
HgTe	3.99	1360	113	230	343	25

a. The U values are in eV, and the others in kg/mm².

Appendix E

ELECTRONIC STRUCTURES OF HgTe AND CdTe SURFACES AND HgTe/CdTe INTERFACES

Electronic structures of HgTe and CdTe surfaces and HgTe/CdTe interfaces

J. T. Schick* and S. M. Bose

Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

A.-B. Chen

Department of Physics, University of Auburn, Auburn, Alabama 36849

(Received 12 September 1988; revised manuscript received 26 June 1989)

A Green's-function method has been used to study the surface and interface electronic structures of the II-VI compounds HgTe and CdTe. Localized surface and resonance states near the cation-terminated (100) surface of CdTe and the anion-terminated surface of HgTe have been found for the ideal surfaces. The energies and strengths of these surface states are altered by surface perturbations. The bulk states near the surface are drastically modified by the creation of the surface, but the band gaps remain unchanged. Numerical evaluation of the local densities of states at the Γ and J points shows that, at the (100) interface of HgTe/CdTe, the previously observed surface states are no longer present. However, in the interface region, bulk states of one material penetrate some distance into the other material.

I. INTRODUCTION

One of the earliest formulations used in the investigation of surfaces and interfaces was based on the linear combination of atomic orbitals (LCAO) method. Goodwin¹ first applied the LCAO model to study the existence conditions for localized Tamm² states in a crystal. He found that these states occur when the diagonal Coulomb integrals and the off-diagonal resonance integrals of the surface are allowed to be different from those of the bulk, for systems of single-level or of *sp*-hybridized atoms. At the same time, Shockley³ investigated a periodic potential that is terminated at its maximum and found that, under the condition that the bulk bands crossed, surface states exist in the middle of the band gap. Shockley states, as they have come to be known, come into being when the perturbations at the surface are small in comparison to the widths of the allowed energy bands. Koustecky⁴ and others⁵ generalized Goodwin's model to study the energy and existence conditions of surface states. Electronic interface states were studied within the LCAO method by Davison and Cheng.⁶ An exact tight-binding solution for the surface and interface problems of a one-dimensional semiconductor was obtained by Dy and co-workers.^{7,8}

Since these early investigations, many other papers have appeared in the literature for the study of surface and interface properties of solids. Kalkstein and Soven (KS) (Ref. 9) introduced a Green's-function (GF) formalism to study the surface electronic properties of semi-infinite crystals. This is a relatively simple but powerful method by which both the bulk and surface properties of a semi-infinite crystal can be studied. The method can be generalized in a straightforward manner to study the electronic properties of an interface formed by joining two semi-infinite crystals. Because of its simplicity and power the method of KS was widely applied in many calculations during the decade following its develop-

ment.¹⁰⁻¹⁴ For the same reasons we apply this method in this work to investigate surface and interface electronic structures of the pure II-VI compounds CdTe and HgTe.^{15,16}

Despite the intense interest in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ over the past three decades due to its application in the preparation of infrared detectors, there have been relatively few experimental and theoretical investigations of the surface and interface properties of this system.¹⁷⁻¹⁹ Since the KS method is well suited for the study of the surface properties of a system described by a tight-binding Hamiltonian, we use here a Hamiltonian closely related to that obtained by Hass, Ehrenreich, and Velicky (HEV) (Ref. 20) for HgTe and CdTe in the empirical tight-binding approximation (ETBA) based on the LCAO interpolation scheme of Slater and Koster.²¹ In the ETBA the predicted band structures of the pure compounds HgTe and CdTe are matched to experimentally determined band energies.²²

In Sec. II, we introduce the formalism of KS for the description of the surface properties of II-VI materials along with the extension of the technique to the problem of interfaces of these materials. Section III contains the results of our calculation and a discussion.

II. FORMALISM

In the formalism of Kalkstein and Soven,⁹ a pair of semi-infinite crystals is formed by introducing a cleavage plane into an infinite crystal in one crystallographic direction. A GF describing the electronic properties of the semi-infinite systems is derived from the GF of the infinite crystal and a scattering potential representing the cleavage. When combined with a tight-binding formalism in which nearest- and next-nearest-neighbor interactions are included, the scattering potential is relatively easy to calculate making application to realistic systems simple. The Hamiltonian is constructed, following HEV,

for HgTe and CdTe using sp^3 basis states and including spin.^{20,23} The parameters used are those of Slater and Koster²¹ as determined by HEV (Ref. 20) except that, in our case, the values of $Esx(110)$ and $Esx(011)$ are interchanged. This produces band structures^{15,16} that are qualitatively similar to those of Bryant which offer improved conduction-band mass.¹⁸ We calculate the GF, G associated with this Hamiltonian by using the defining equation

$$G = (E + i\lambda - H)^{-1}, \quad (1)$$

where H is the Hamiltonian of the infinite unperturbed system. It is understood that G is to be calculated in the limit as the positive imaginary part λ approaches zero. These calculations are performed in k space, utilizing fully the periodicity of the crystal.

For the surface (interface) calculations, since we no longer have translational periodicity along the direction perpendicular to the surface (interface), we cannot use an ordinary k -space representation. Kalkstein and Soven⁹ assume periodic structure parallel to the surface and use a representation consisting of states which are localized on planes of atomic sites parallel to the surface and represented by the index n , and of Bloch states reflecting the periodic symmetry within the planes which are represented by the index k_{\parallel} . This is the mixed or Bloch-Wannier representation. We assume the same type of symmetry in the interface system. The Hamiltonian and the GF of the bulk crystal as well as the Hamiltonians of the semi-infinite and interface systems must all be expressed in this mixed representation. The formulas for the semi-infinite system were derived by KS and are the same for us if we reinterpret the algebraic expressions as matrix equations for the sp^3 -spin basis states. Note that in this paper we examine the surfaces and interfaces perpendicular to the [100] direction for pure CdTe and HgTe samples. For these compounds, this structure corresponds to an arrangement in which the anions and cations are placed in alternating planes parallel to the surface or interface. For notational simplicity, in the following, a cation-anion pair of planes is given a single layer index n , with the species index v left to distinguish between the two species (layers).^{15,16}

For basis states of the infinite system, KS used states analogous to $|\alpha\sigma\nu\mathbf{k}\rangle$, where $\alpha=s,x,y,z$ indicates the atomic orbital, the spin (\uparrow, \downarrow) is represented by σ , $\nu=a$ or c stands for anions or cations, respectively, and \mathbf{k} is the wave vector. Following KS, we write the Bloch-Wannier states for our system as

$$|\alpha\sigma\nu\mathbf{k}_1\rangle = \frac{1}{N_1} \sum_{\mathbf{k}_1} e^{-i\mathbf{k}_1 \cdot \mathbf{R}_{n\nu}} |\alpha\sigma\nu\mathbf{k}\rangle, \quad (2)$$

where $\mathbf{R}_{n\nu}$ is the position of the ion sublattice represented by ν on the plane labeled by n , and \mathbf{k}_{\parallel} and \mathbf{k}_1 are the components of \mathbf{k} , parallel and perpendicular to the surface (interface) plane, respectively. This basis set reflects the symmetry of the semi-infinite system and is therefore well suited for our purpose. In the following we suppress the \mathbf{k}_1 index for compactness of notation, as in

$$G(n, n') = \sum_{\alpha\sigma\nu} \sum_{\alpha'\sigma'\nu'} \langle \alpha\sigma\nu\mathbf{k}_{\parallel} | G | \alpha'\sigma'\nu'\mathbf{k}_{\parallel} \rangle \times |\alpha\sigma\nu\mathbf{k}_{\parallel}\rangle \langle \alpha'\sigma'\nu'\mathbf{k}_{\parallel}|, \quad (3)$$

which is the GF submatrix between layer n and layer n' .

From the site-diagonal GF, $G(n, n)$, we may calculate the local density of states

$$\rho_{n\nu}(E) = -(1/\pi) \text{Im Tr}_\nu G(n, n), \quad (4)$$

where Tr_ν indicates that the trace is to be carried out only for the species (layer) given by ν .

The GF of the semi-infinite crystal, G' is found through the application of the Dyson equation,

$$G' = G + GVG'. \quad (5)$$

Besides calculating the matrix elements of the infinite crystal GF, G , we must also find the matrix elements of the scattering potential, V , introduced by the creation of the surface. We label the double layers in the crystal with integers such that the cleavage plane passes between the double layers -1 and 0 , and the semi-infinite crystal of interest starts at the zeroth layer and has layer indices $n \geq 0$. Because $G'(n, n')$ is zero for all n and n' less than zero, the only matrix elements of the scattering potential that enter into Eq. (5) in the present model are $V(-1, 0)$ and $V(0, -1)$ which express the severing of the interactions across the cleavage plane. We may also include a diagonal term $V(0, 0)$, if we wish, to allow for surface perturbations such as relaxation and environmental shifts, into our calculation. These scattering potentials have explicit dependence on \mathbf{k}_{\parallel} and the pair of indices associated with V refer to the double layers involved in the interaction. In terms of the Hamiltonian matrices, the scattering potential matrix is given by

$$V = H' - H, \quad (6)$$

where H' is the Hamiltonian of the semi-infinite crystal. The scattering potential matrix elements describing the breaking of interactions across the cleavage plane are⁹

$$\begin{aligned} V(0, -1) &= -H(0, -1), \\ V(-1, 0) &= -H(-1, 0), \end{aligned} \quad (7)$$

where, as before, the \mathbf{k}_1 index has been suppressed for compactness of notation. The matrix element $V(0, 0)$ is introduced parametrically to account for the shifts in the atomic levels and hopping interaction at the surface layer.

Explicitly, a general matrix of G' is found from Eq. (5) as

$$\begin{aligned} G'(n, n') &= G(n, n') \\ &+ [G(n, 0)V(0, 0) \\ &+ G(n, -1)V(-1, 0)]G'(0, n'), \end{aligned} \quad (8)$$

with n and $n' \geq 0$. To solve this equation, it is first necessary to solve for $G'(0, n')$ which is done by setting n equal to zero in Eq. (8) and solving for the matrix element

$$G'(0, n') = [1 - G(0, 0)V(0, 0) - G(0, -1)V(-1, 0)]^{-1} G(0, n'). \quad (9)$$

To find the general matrix element of G' we need only to substitute Eq. (9) into Eq. (8). To include environmental effects on the surface atoms, we introduce a parameter δ which measures a shift of the surface layer on-site energies. Also, we introduce a parameter d which represents the strength of the interactions between the first two planes of atoms in the semi-infinite crystal relative to the same interactions in the bulk. The parameters δ and d are included in the matrices $V(0, 0)$ and $V(-1, -1)$.^{15,16}

This model was first extended to the interface between two model semi-infinite one-band crystals by Yaniv.¹³ We further extend the technique to real crystals using the sp^3 with spin basis states. The interface is formed between crystal A ($n \leq -1$) and crystal B ($n \geq 0$). In joining the crystals, the interactions between the Te and Cd planes, as well as those between the Te planes, across the interface, are taken to be the same as in the bulk. Since there are no data available for the hopping integrals between Hg and Cd planes we assume them to be equal to the average of the interactions between bulk-crystal (100) Cd planes in CdTe and interactions between bulk-crystal (100) Hg planes in HgTe. The interactions between the A and B sides of the interface are included in the scattering potentials $V'(0, -1)$ and $V'(-1, 0)$. To our Green's function we add subscripts α and β which take on the values A and B to indicate the side of the interface involved in the calculations. Once again solving the Dyson equation for the interface GF, g (Ref. 16),

$$g = G' + G'V'g$$

we find the interface GF matrix elements as

$$g_{\alpha\beta}(m, n) = G'_{\alpha}(m, n)\delta_{\alpha\beta} + G'_{\beta}(m, 0)V'(0, -1)g_{A\beta}(-1, n)\delta_{\alpha\beta} + G'_{\alpha}(m, -1)V'(-1, 0)g_{B\beta}(0, n)\delta_{\alpha\beta}, \quad (10)$$

where

$$g_{AA}(-1, n) = [1 - G'_A(-1, -1)V(-1, 0)G'_B(0, 0) \times V(0, -1)]^{-1} G'_A(-1, n), \quad (11a)$$

$$g_{BB}(0, n) = [1 - G'_B(0, 0)V(0, -1)G'_A(-1, -1) \times V(-1, 0)]^{-1} G'_B(0, n), \quad (11b)$$

$$g_{BA}(0, n) = G'_B(0, 0)V(0, -1)g_{AA}(-1, n), \quad (11c)$$

$$g_{AB}(-1, n) = G'_A(-1, -1)V(-1, 0)g_{BB}(0, n), \quad (11d)$$

and $\delta_{\alpha\beta}$ is the Kronecker delta. We have assumed a perfect match in the lattice spacing across the interface which is nearly true for HgTe and CdTe with a difference in spacing of only 0.3%.

In the evaluation of the GF of the bulk system a great deal of numerical integration is required. The integrals are evaluated through the use of a Lagrange interpolation scheme in which the time to calculate the band structures is reduced by evaluating them exactly at only a few points

in k space and interpolating for the intermediate values.²⁴ Accuracy is ensured through sampling the functions at a sufficient number of points. All calculations are performed with a small positive imaginary component in the energy and the final results at the real energy axis are evaluated through the use of the analytic continuation procedure of Hass, Velicky, and Ehrenreich.²⁵

One quite useful feature of our calculation is that the evaluation of the GF's of the semi-infinite crystal and the interface requires much less computational effort than the evaluation of the infinite crystal GF.¹⁶ As a result of this speed, we are able to consider several values of the environmental shift and surface-nearest-neighbor interaction. In order to observe the effects of the variation of these parameters on the localized states we examine the local densities of states (LDOS) at various values of the parameters for a few CdTe and HgTe surfaces and interfaces. These LDOS are evaluated at fixed values of k , so that we may find the positions of the localized state bands in the surface Brillouin zone. For details of the evaluation of the matrix elements of the infinite crystal GF and those of the scattering potentials required for our calculation, one is referred to Refs. 15 and 16.

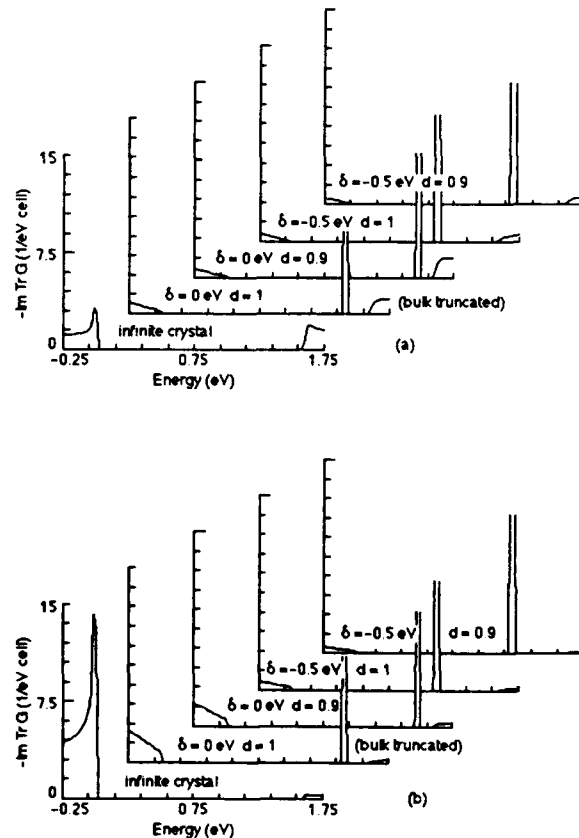


FIG. 1. Dependence of the LDOS on the parameters δ and d as compared to the infinite crystal for (a) the cation surface layer and (b) the first interior anion layer at Γ ($k_{\parallel} = 0$) for CdTe cation-terminated (100) surface.

III. RESULTS AND DISCUSSION

Once the matrix elements of the infinite crystal GF and the relevant scattering potentials are specified, one can calculate the matrix elements of the GF of the semi-infinite sample and that of the sample with the interface using Eqs. (8) and (10), respectively. One can further calculate the LDOS at various layers using an equation of the type of Eq. (4). In this paper we present the LDOS calculated at the symmetry points Γ and J and at energies near the valence- and the conduction-band edges of the sample where the principal band gaps occur.

In Figs. 1 and 2 we plot the surface LDOS at the Γ point for various combinations of the environmental shift parameter δ and the geometric shift parameter (relaxation parameter) d for CdTe cation-terminated and HgTe anion-terminated crystals, respectively. We begin by noting, in Fig. 1, the existence of a localized surface state (bound state) which is the sharp structure that appears in the LDOS within the band gap of the CdTe, whether or

not there is a surface perturbation. The ideal cation-truncated (100) surface of CdTe has a surface state at Γ with energy lying at $E \approx 0.6$ eV above the top of the valence band. We also see that the bulk densities of states are modified as a result of the introduction of a surface but the band gap of CdTe remains unchanged. We note that the position of the bound state is sensitive to

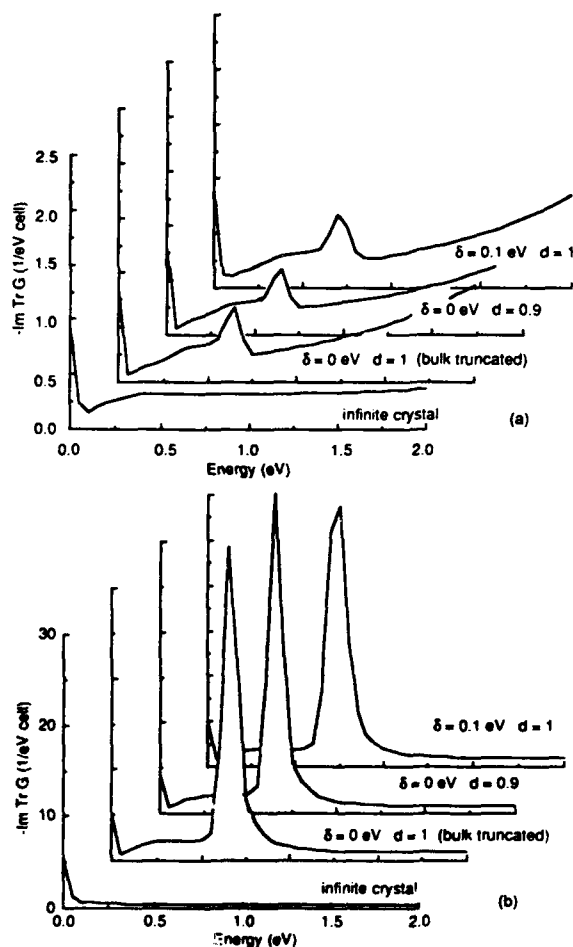


FIG. 2. Dependence of the LDOS on the parameters δ and d as compared to the infinite crystal for (a) the anion surface layer and (b) the first interior cation layer at Γ ($k_{\parallel} = 0$) for HgTe cation-terminated (100) surface.

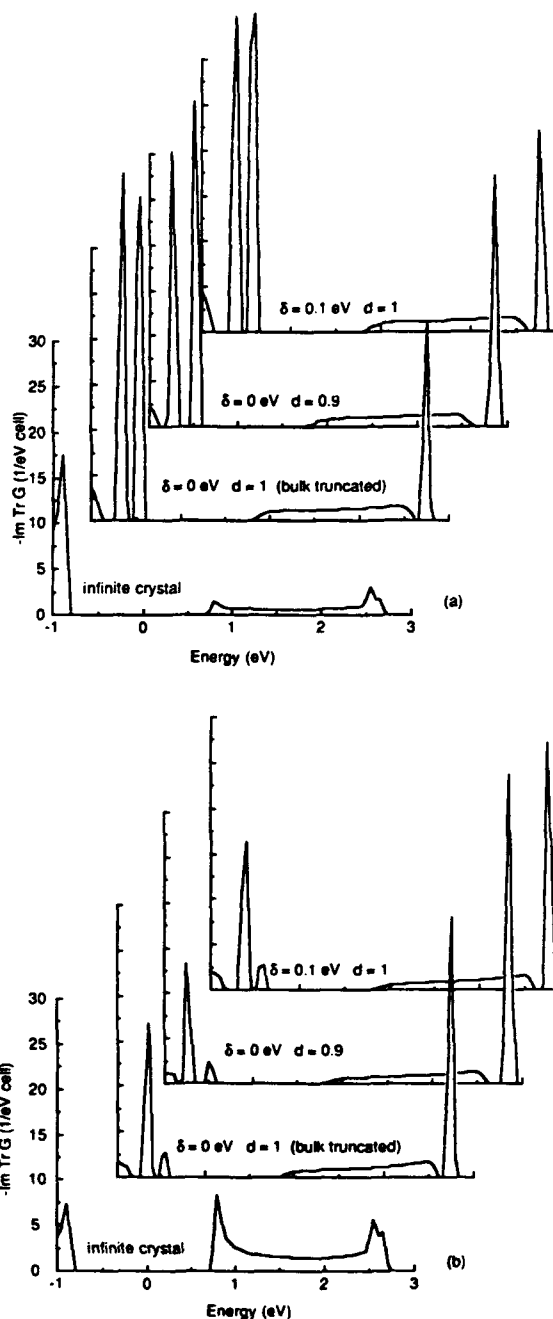


FIG. 3. Dependence of the LDOS on the parameters δ and d as compared to the infinite crystal for (a) the anion surface layer and (b) the first interior cation layer at J [$k_{\parallel} = \pi/a(\sqrt{2}, \sqrt{2})$] for HgTe anion-terminated (100) surface.

variations in both d and δ . In Fig. 2, we notice that the surface has again introduced a marked change in the LDOS of HgTe. A resonance state appears in the form of a spike inside the conduction band of semimetal HgTe. The contribution of this resonance to the LDOS is strongest for anion layers indicating that anion states are the most likely constituents of this state. Similar surface states have been calculated by Bryant¹⁸ for the special case where no surface perturbations exist.

The surface states calculated for the HgTe at the J point where there exists a gap are shown in Fig. 3. These states exhibit the same dependence on the interactions that was seen at the Γ point. Here instead of a single peak we see a pair of bound states just above the valence band within the band gap. Also above the conduction band there is a bound state that was not observed for the Γ point. Anions contribute more strongly to the bound states inside the band gap while the cation contribution is stronger to the bound state above the conduction band. Also notice that the bulk states are strongly modified by

the introduction of the surface and that the band gaps are not altered by the surface.

Our calculation of the layer dependence of the surface states shows that these states become progressively less significant as we examine deeper inside the crystal indicating these states are indeed localized bound states. The bulk LDOS in the deeper layers, on the other hand, start resembling the infinite crystal LDOS.

As mentioned before, the interface calculations have been performed assuming no lattice mismatch in the crystal structures of the HgTe and CdTe compounds, thus neglecting all strains that may be present at the interface. The parameters chosen for our calculation also do not allow for valence-band offset. With these assumptions, dramatic effects are seen at the interface of HgTe and CdTe as shown in Figs. 4 and 5, where the LDOS in the CdTe side of the interface are presented at Γ and J points, respectively. The most obvious effect is the lack of localized states that were previously seen at the free surfaces of these materials. Instead, we find that, close to the interface, the bulk states of HgTe that lie in the region between 0 and 1.6 eV appear in the band gap of

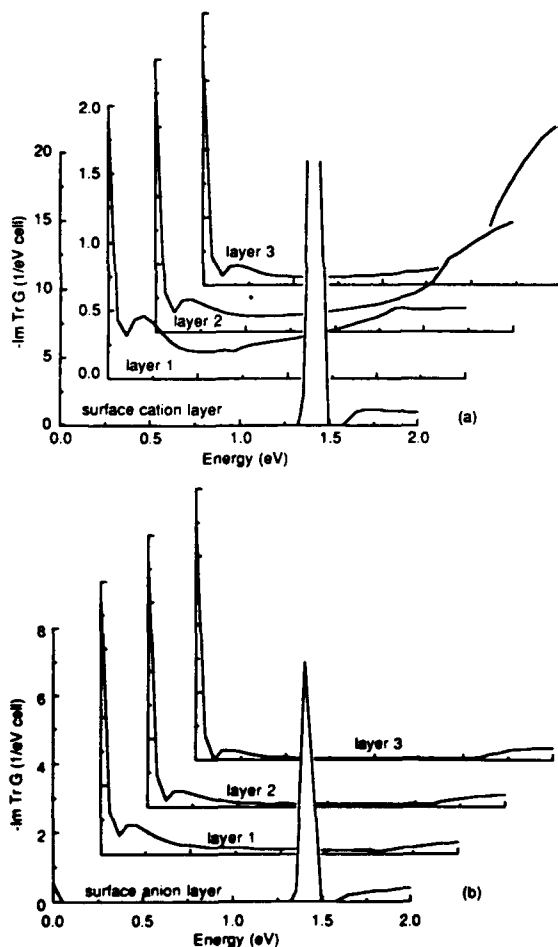


FIG. 4. Comparison of the surface LDOS of the cation-terminated (100) surface of CdTe with the LDOS plotted (a) at successive cation layers and (b) anion layers for the CdTe side of the (100) interface CdTe/HgTe at Γ .

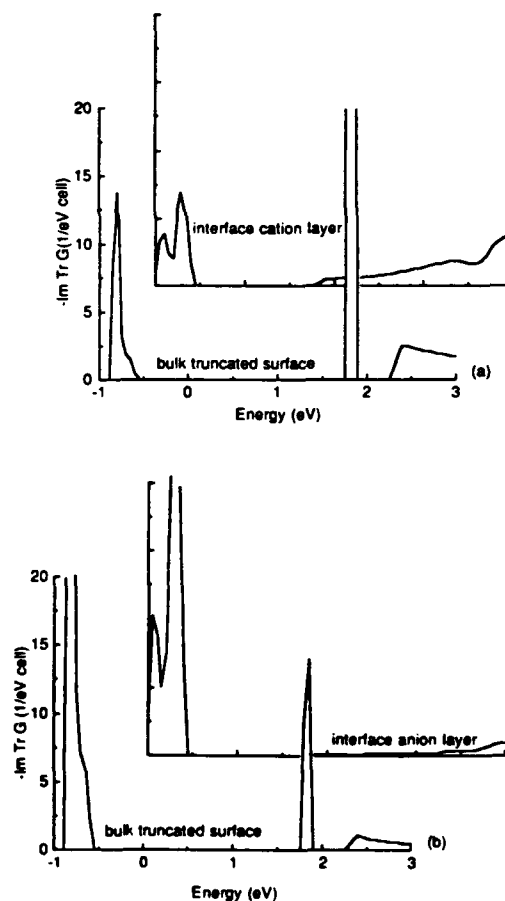


FIG. 5. Comparison of the surface LDOS of the cation-terminated (100) surface of CdTe with the LDOS plotted (a) at first cation layer at the interface and (b) the first interior anion layer for the CdTe side of the (100) interface CdTe/HgTe at J .

CdTe at the Γ point. These states in the CdTe gap become less significant at layers further from the interface while at the same time the LDOS in the energy ranges of the conduction and valence bands of CdTe increase in strength. Thus we have a narrowing of the gap of CdTe in a limited region near the interface which allows conduction electrons to penetrate into the CdTe from the HgTe over a short distance. In Fig. 5, we explicitly see how the gap at the CdTe layers is reduced with respect to pure CdTe at J . Calculation of the LDOS in the HgTe side of the interface shows similar effects.

The above results describe how the LDOS change from the bulk to the surface and then from the surface to the interface. While the bulk and the interface results can be considered realistic, the surface results may not be, since the surface reconstruction has not been included in the calculation. Recent experiments^{17,26-28} and a structural theory²⁹ indicate that the surfaces of CdTe and some other II-VI compounds undergo reconstructions similar to those on the surfaces of III-V compounds. However, similar measurements are yet to be made on HgTe surfaces and CdTe/HgTe interfaces. Our calculations indicate that changes in the electronic structure in the form of environmental shifts at the surface lead to only minor changes in the localized surface and the bulk LDOS. Whether a surface reconstruction will result in substantial modification of these states is yet to be determined.

In conclusion, we have described in this paper how one can incorporate the GF method of KS to calculate the surface and interface structures of II-VI compounds

without having to deal with the artificial super-cell approximation. Our calculation shows that drastic changes in the LDOS can occur because of creation of surfaces and interfaces. Localized surface or resonance states appear above the top of the valence band, and the effects of the surface on the LDOS persist at several layers inside the bulk. At the interface, there are no localized states but the bulklike states related to one material penetrate into several layers inside the other material resulting in narrowing of band gaps close to the interface for the wide band-gap material. At present there are no systematic experimental data available for comparison with our conclusions.

Finally, even though we have not included all aspects necessary for a complete description of the surfaces and interfaces, our work can be considered to be the first step toward the understanding of the surface and interface electronic structures of the II-VI compounds. To our knowledge, the results presented here are new for the interface and more general than any previous surface calculations on the II-VI compounds. Since the method is flexible, it will be possible to incorporate realistic structural models dealing with the surface reconstructions, when they become available, to calculate more realistic electronic properties of such surfaces and interfaces.

ACKNOWLEDGMENTS

This work was supported by National Aeronautics and Space Administration Grant No. NAG 1708.

*Present address: Department of Physics and Astronomy, Wayne State University, Detroit, MI 48202.

¹E. T. Goodwin, Proc. Cambridge Philos. Soc. **35**, 232 (1939).

²I. Tamm, Z. Phys. **76**, 849 (1932); Phys. Z. Sowjet. **1**, 733 (1932).

³W. Shockley, Phys. Rev. **56**, 317 (1939).

⁴J. Koutecky, Phys. Rev. **108**, 13 (1957).

⁵J. Alstrup, Surf. Sci. **20**, 325 (1970).

⁶S. G. Davison and Y. C. Cheng, Int. J. Quantum Chem. Symp. **2**, 303 (1968).

⁷K. S. Dy, S.-Y. Wu, and T. L. Spratlin, Phys. Rev. B **20**, 4237 (1979).

⁸J. D. Brasher and K. S. Dy, Phys. Rev. B **22**, 4868 (1980).

⁹D. Kalkstein and P. Soven, Surf. Sci. **26**, 85 (1971).

¹⁰E. N. Foo and H. S. Wong, Phys. Rev. B **9**, 1857 (1974); **10**, 4819 (1974).

¹¹S. M. Bose and E. N. Foo, Phys. Rev. B **10**, 3534 (1974).

¹²S. M. Bose and J. M. Yuan, Phys. Rev. B **24**, 1934 (1981).

¹³A. Yaniv, Phys. Rev. B **17**, 3904 (1979); **22**, 4776 (1980).

¹⁴N. F. Berk, Surf. Sci. **48**, 289 (1975).

¹⁵J. T. Schick and S. M. Bose, Prog. Surf. Sci. **25**, 107 (1987).

¹⁶J. T. Schick, Ph. D. thesis, Drexel University, 1987 (unpublished).

¹⁷W. E. Spicer, J. A. Silverman, P. Morgan, I. Lindau, and J. A. Wilson, J. Vac. Sci. Technol. **21**, 149 (1982).

¹⁸G. W. Bryant, Phys. Rev. Lett. **55**, 1786 (1985); Phys. Rev. B **35**, 5547 (1987); **36**, 6465 (1987).

¹⁹J. N. Schulman and T. C. McGill, Appl. Phys. Lett. **34**, 663 (1979).

²⁰K. C. Hass, E. Ehrenreich, and B. Velicky, Phys. Rev. B **27**, 1088 (1983).

²¹J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

²²D. J. Chadi, J. P. Walter, M. L. Cohen, V. Petroff, and M. Balkanski, Phys. Rev. B **5**, 3058 (1972); A. Moritani, K. Taniguchi, C. Hamaguchi, and J. Nakai, J. Phys. Soc. Jpn. **34**, 79 (1973); N. J. Schevchik, J. Tejada, M. Cardona, and D. W. Langer, Phys. Status Solidi B **59**, 87 (1973); L. Ley, R. A. Pollak, F. R. McFeely, S. P. Kowalczyk, and D. A. Shirley, Phys. Rev. B **9**, 600 (1974).

²³D. J. Chadi, Phys. Rev. B **16**, 790 (1977).

²⁴A.-B. Chen, Phys. Rev. B **16**, 3291 (1977).

²⁵K. C. Hass, B. Velicky, and H. Ehrenreich, Phys. Rev. B **29**, 3697 (1984).

²⁶P. John, T. Miller, T. C. Hsieh, A. P. Shapiro, A. L. Wachs, and T.-C. Chiang, Phys. Rev. B **34**, 6706 (1986).

²⁷K. O. Magnusson, S. A. Foldstrom, and P. E. S. Person, Phys. Rev. B **38**, 5384 (1988).

²⁸P. Lu and D. J. Smith, Phys. Rev. Lett. **59**, 2177 (1987).

²⁹Y. R. Wang and C. B. Duke, Phys. Rev. B **36**, 2763 (1987).

Appendix F
ALLOY STATISTICS AND PHASE DIAGRAMS

IV ALLOY STATISTICS AND PHASE DIAGRAMS

4.1 Mixing Free Energy, Miscibility Gap and Order-Disorder Transitions

Consider an ideal pseudobinary semiconductor alloy $A_{1-x}B_xC$ in the zincblende structure in which the alloy atoms A and B randomly occupy their fcc sublattice sites while the C atoms occupy the other fcc sublattice. If the C atoms are treated as spectators, then this pseudobinary alloy behaves like a fcc binary alloy $A_{1-x}B_x$ in the statistical mechanics formalism. However, as evidenced by the EXAFS experiment discussed in Chapter I, the atomic positions in a real semiconductor alloy are distorted slightly from the zincblende sites. Because the bonding is covalent the C atoms mediate the energies and in that sense affect the statistics. It will become clear later that the distribution of A and B atoms in an alloy is never completely random. This has important consequences on many physical properties. For example, a knowledge of structural energies and associated atomic distributions is essential to accurate calculations of phase diagrams.

In this chapter we will be dealing with equilibrium statistics and phase diagrams. The equilibrium state at a fixed temperature T and pressure P is the one with the minimum Gibbs free energy. The equilibrium state may contain only one phase, for example, an ordered alloy in a given crystal structure, or a disordered solution with a uniform concentration. It may also contain several phases, for example, an ordered alloy plus a disordered one, or two disordered phases with different concentrations. Gibbs' phase rule (e.g., Landau and Lifshitz, 1986) states that a solution containing n species can have up to a maximum of $n + 2$ phases coexisting in equilibrium. Thus to determine the equilibrium state at constant temperature and pressure of an alloy requires knowledge of the Gibbs free energies of all possible phases.

Under normal pressures of the order of one atmosphere, the difference between the Gibbs free energy G and the Helmholtz free energy F , $G - F = PV$, is insignificant for a solid or liquid. Thus, it is sufficient to use F for most cases. For a disordered zincblende pseudobinary alloy $A_{1-x}B_xC$, it is

convenient to define a mixing free energy ΔF as a function of x and T at a fixed pressure

$$\Delta F(x,T) = F(x,T) - (1-x)F_{AC}(T) - xF_{BC}(T), \quad (4.1.1)$$

where F , F_{AC} and F_{BC} are respectively the Helmholtz free energies for the alloy, pure AC and pure BC compounds containing the same number of C atoms. ΔF can be written as

$$\Delta F = \Delta E - T\Delta S, \quad (4.1.2)$$

where ΔE is the mixing energy and ΔS is the mixing entropy defined similarly to ΔF . Again, because the magnitude of $\Delta(PV)$ is small, ΔE and the mixing enthalpy ΔH are used interchangeably. If $\Delta E > 0$ at $T = 0$, which turns out to be the case for most semiconductor alloys, then ΔF as a function of x at different fixed T will have the schematic shapes of the curves shown in Fig. 4.1.1a. The curve labelled $T_3 > T_c$ represents a typical high temperature curve. It tells us that disordered alloys with any concentration fraction x are stable at this temperature. However, the low-temperature curve labelled $T_2 < T_c$ indicates there is a miscibility gap for x within the interval $x_1 < x < x_2$. Here x_1 and x_2 are the points at which the common tangent line touches the ΔF curve. This curve means that the alloy is thermally stable against decomposition for $x \leq x_1$ and $x \geq x_2$ but not inside the interval $x_1 < x < x_2$. For x inside the gap, the alloy tends to decompose into two alloys with concentrations x_1 and x_2 with proportions P_1 and P_2 governed by the lever rule, $P_1 = (x_2 - x)/(x_2 - x_1)$ and $P_2 = (x - x_1)/(x_2 - x_1)$, because this decomposition minimizes the free energy. There are two other special concentration values x'_1 and x'_2 called the spinodal points. These occur at inflection points of ΔF at each $T < T_c$, i.e., $\partial^2 \Delta F / \partial x^2 = 0$ at these x values. These values separate x space into unstable and metastable regions. For $x_1 < x < x'_1$ or $x'_2 < x < x_2$, the alloy is metastable against local decomposition, because the ΔF value for any x in these regions is lower than the lever rule average value of ΔF when the alloy begins to separate into domains with compositions in the neighborhood of x . This causes a barrier to the alloy decomposition into its equilibrium state. An alloy with $x'_1 < x < x'_2$ is inherently unstable because there is no decomposition barrier. We can construct the binodal

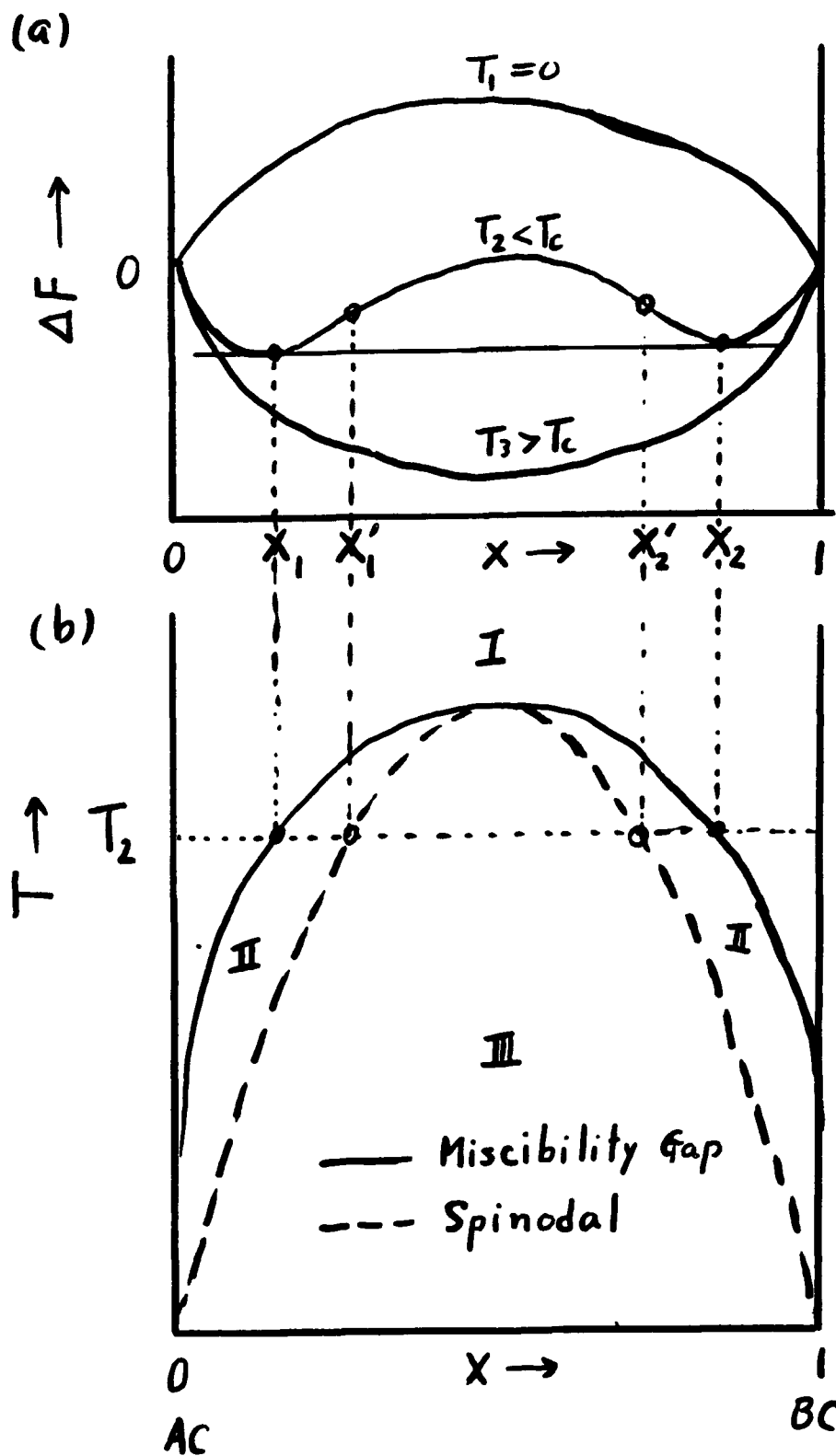


Figure 4.1.1 (a) Schematic picture of mixing free energies ΔF as a function of composition x of a pseudobinary alloy $A_{1-x}B_xC$ with a positive mixing enthalpy at three temperatures $T_1 = 0$, $T_2 < T_c$ and $T_3 > T_c$. (b) Schematic miscibility gap and spinodal curves. Regions I, II, and III are stable, metastable and unstable respectively.

and spinodal curves by continuously varying the temperatures and tracing the values of these gaps, as shown in Fig. 4.1.1b. These gap values become closer to each other as the temperature increases until a critical temperature T_c is reached, where all the values x_1 , x_2 , x'_1 and x'_2 merge into one value. Beyond T_c , the disordered phase is stable for all alloy concentrations.

If $\Delta E < 0$ at $T = 0$, then there is a tendency to form a long-range ordered alloy. Whether the alloy is ordered or disordered depends upon the temperature. Figure 4.1.2 schematically compares the mixing free energy of the disordered alloy against that of an ordered compound ABC_2 . With a negative mixing enthalpy, the energy of the ordered system ΔE_0 is lower than the disordered alloy, i.e., $\Delta E_0 < \Delta E$ at $T = 0$, so that the ordered state is the equilibrium state. As temperature increases, ΔF of the disordered alloy becomes more negative, mainly because the entropy term in Eq. (4.1.2) decreases roughly linearly in T . The system remains in the ordered phase until a transition temperature T_0 is reached when ΔF is equal to ΔE_0 . The disordered alloy then becomes the stable state for temperatures greater than T_0 .

4.2 Analytical Models

Since pseudobinary alloys $A_{1-x}B_xC$ statistically are similar to binary alloys $A_{1-x}B_x$, the binary results will be used whenever they do not cause any confusion. The statistical theory for binary solid solutions is mathematically equivalent to the three-dimensional Ising model in magnetism. Even for the simplest case with nearest-pair interactions, an analytical solution of this model remains one of the most challenging problems in theoretical physics. However, systematic computational methods are available, as will be discussed later. This section briefly reviews several analytical models, which are useful for illustrating the basic concepts and for semi-empirical phase diagram evaluation. This review will focus on disordered alloys with positive mixing energies, because this is the most relevant case for most bulk semiconductors. However, results related to ordered cases will also be discussed. More detailed examples of spontaneous ordering will be considered in Chapter 7.

4.1.2

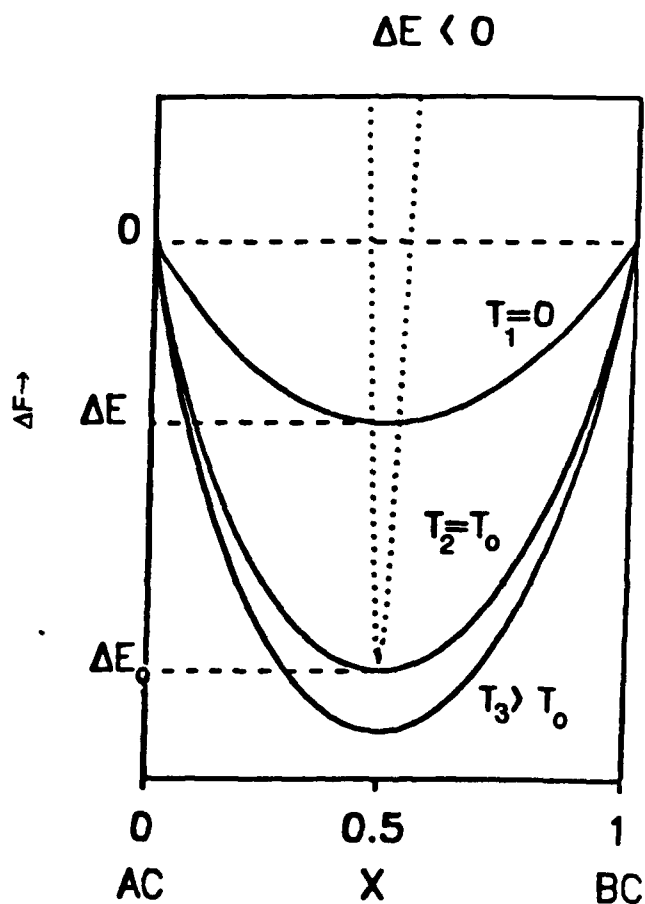


Figure 4.1.2 Schematic picture of mixing free energies ΔF as a function of alloy composition x of a pseudobinary alloy $A_{1-x}B_xC$ with a negative mixing enthalpy at three temperatures $T_1 = 0$, $T_2 = T_c$ and $T_3 > T_c$. The dotted curve is for an ABC_2 ordered phase.

Ideal-solution Model

This model is equivalent to the case where the mixing energy vanishes, $\Delta E = 0$ in Eq. (4.1.2). Thus in this approximation, the mixing free energy ΔF results totally from the additional entropy arising from a random arrangement of N_A A-atoms and N_B B-atoms on their N lattice sites with $N = N_A + N_B$ and $x = N_B/N$. The number of different arrangements is

$$\Phi_0 = \frac{N!}{N_A! N_B!} \quad (4.2.1)$$

The corresponding mixing entropy is $\Delta S = k \ln \Phi_0$. Using the Stirling approximation $\ln N! \cong N \ln N - N$, the mixing entropy for the random alloy is obtained,

$$\Delta S = -Nk[(1-x) \ln(1-x) + x \ln x] \quad (4.2.2)$$

This model always yields negative values of $\Delta F = -T\Delta S$ for all x and T . As a function of x , ΔF in Eq. (4.2.2) is always concave upward, so no miscibility gap exists at any T .

ZEROTH APPROXIMATION

When ΔE is not zero but the mixing entropy is still set equal to the random alloy result, the model was named by Guggenheim (1952) the "strict-regular solution" model, or the zeroth approximation. If, following shapes often experimentally observed in pseudobinary alloys, ΔE is assumed to have an x dependence given by

$$\Delta E = vx(1-x)\Omega \quad (4.2.3)$$

then miscibility gaps exist at low temperature for positive molar interaction parameters Ω . In Eq. (4.2.3), v is the molar number $v = N/N_0$ with N_0 being Avogadro's number.

Equation (4.2.3) as we shall now demonstrate, is the expression for the mixing energy if pair-wise interaction energies and a random distribution are assumed. Let z be the coordination number. Note that z is the number of nearest neighbor atoms surrounding a site in the binary case, but it

represents the number of the second-neighbor atoms in a pseudo-binary alloy. For example, the value of z for a zincblende alloy is 12. Then the total number of pairs involving the A and B atoms is $M = zN/2$. Let M_{AA} , M_{AB} and M_{BB} be the number of AA, AB, and BB pairs respectively, and let ϵ_{AA} , ϵ_{AB} and ϵ_{BB} be the corresponding pair interaction energies. We further define the fractions $r_{ij} = M_{ij}/M$. Not all these pair fractions are independent. Let $r = r_{AB}$, then $r_{AA} = 1 - x - 0.5r$ and $r_{BB} = x - 0.5r$. The mixing energy is governed only by the AB pair fraction r ,

$$\Delta E = M_{AA}\epsilon_{AA} + M_{AB}\epsilon_{AB} + M_{BB}\epsilon_{BB} - (1-x)M\epsilon_{AA} - xM\epsilon_{BB} = Mr\epsilon \quad (4.2.4)$$

where ϵ is the excess energy per pair and is defined as

$$\epsilon = \epsilon_{AB} - (\epsilon_{AA} + \epsilon_{BB})/2. \quad (4.2.5)$$

For a random alloy the pair probability is given by

$$r = 2(1-x)x, \quad (4.2.6)$$

so the mixing energy has the form of Eq. (4.2.3) with the interaction parameter given by

$$\Omega = N_0 z \epsilon. \quad (4.2.7)$$

This strict-regular model is the simplest one that contains some aspect of reality, because it relates a positive mixing enthalpy to a miscibility gap and critical temperature T_c . For example, T_c can be obtained explicitly by setting $\partial^2 \Delta F / \partial x^2$ equal to zero at $x = 1/2$,** where $\Delta F = \Delta E - T_c \Delta S$ with E in the form of Eq. (4.2.3) and ΔS having the random alloy expression in Eq. (4.2.2). The result is

$$T_c = \frac{\Omega}{2R}. \quad (4.2.8)$$

where $R = N_0 k$ is the universal gas constant.

* ΔF has its second derivative vanish at $x = 1/2$ in this case because of the assumed symmetric forms of Eqs. (2.2.2) and (2.2.3). In more general cases that extend beyond pair interactions as we shall see, the second derivative vanishes at x values of the composition $x \neq 1/2$.

First Approximation - the Quasi-Chemical Approximation (QCA)

In the zeroth approximation, an energy term is added, but the distribution of A and B atoms is still constrained to be random, a condition that is not consistent with the non-zero excess pair energy ϵ . Intuitively one expects that the pair probability r for the AB pair should be smaller than the random value in Eq. (4.2.6), if ϵ is positive; and vice versa. The first approximation in Guggenheim's notation (1952) corrects this flaw in an approximate way. The correction comes from the mixing entropy. For a random alloy, the number of ways of configuring the A and B atoms, Φ_0 , is given by Eq. (4.2.1). For a specified set of M_{AA} , M_{AB} ; and M_{BB} , the number of ways to arrange A and B atoms should be just Φ_0 times the probability of having the specified set of numbers of pairs (Sher et al. 1987)

$$\Phi = \frac{N!}{N_A! N_B!} \left[\frac{M!}{M_{AA}! M_{AB}! M_{BB}!} y^{2M_{AA}} (2xy)^{M_{AB}} x^{2M_{BB}} \right]. \quad (4.2.9)$$

In Eq. (4.2.9) $y = 1-x$ is the fractional concentration of A atoms. The term outside the bracket is the total distinguishable number of ways to arrange N_A and N_B atoms on N sites, and the bracket is the fraction of these ways, i.e., the probability, M_{AA} , M_{AB} , and M_{BB} pairs will be found when the concentration of B atoms is χ_0 . This probability is the number of ways of arranging the M_{AA} , M_{AB} , and M_{BB} pairs on M pair sites times the probability of M_{AA} pairs, i.e., $(x^2)^{M_{AA}}$, M_{AB} pairs, and M_{BB} pairs. This expression still neglects the correlations between pairs but we shall show in Section 4.10 that in fact it does not.. Equation (4.2.9) should be compared with Guggenheim's (1952) combinatorial formula

$$\Phi_G = \frac{N!}{N_A! N_B!} \frac{(Mx^2)!(Mxy)!(Mxy)!(My^2)!}{M_{AA}! M_{AB}! M_{BA}! M_{BB}!}. \quad (4.2.10)$$

Note that the leading terms of $\ln\Phi$ and $\ln\Phi_G$ in the Stirling approximation are the same, so these two expressions give the same mixing entropy.

However, the physics in Eq. (4.2.9) is more transparent and it is an easier form to extend beyond the pair approximation.

The mixing free energy with Φ given by Eq. (4.2.9) and excess energy ϵ given by Eq. (4.2.5) is

$$\Delta F = Mr\epsilon - kT \ln \Phi_0 + MkT [(x-r/2) \ln(x-r/2) + (y-r/2) \ln(y-r/2) + r \ln(r/2) - (x \ln x + y \ln y)] \quad (4.2.11)$$

The value of r for an equilibrium distribution is the one that minimizes ΔF . Taking the partial derivative $\partial \Delta F / \partial r$ and setting it equal to zero leads to a quadratic algebraic equation for r . The proper solution is

$$r = \frac{4xy}{1 + \left\{ 1 + 4xy \left[\exp \left(\frac{2\epsilon}{kT} \right) - 1 \right] \right\}^{1/2}} \quad (4.2.12)$$

This expression reduces to the correct limits. As ϵ approaches 0, r approaches $2xy$, which is the random limit given in Eq. (4.2.6). For large $\epsilon \gg kT$, r decreases exponentially, $r \propto e^{-\epsilon/kT}$, as expected. If ϵ is negative, then $r > 2xy$, and for large enough $|\epsilon| \gg kT$ the system is eventually driven into compound formation, i.e., $r \rightarrow 2x$ for $x < 1/2$, or $r \rightarrow 2y$ for $x > 1/2$.

The AB pair probability r in Eq. (4.2.12) can also be obtained by considering the following chemical equilibrium



According to the law of mass action, the equilibrium pair fractions obey the following relation

$$\frac{r_{AB}^2}{r_{AA}r_{BB}} = \frac{z_{AB}^2}{z_{AA}z_{BB}} = \frac{(2e^{-\epsilon_{AB}\beta})^2}{e^{-\epsilon_{AA}\beta} e^{-\epsilon_{BB}\beta}} = \frac{4}{e^{2\epsilon\beta}} \quad (4.2.14)$$

where $\beta = 1/(kT)$ and the z_{ij} s are the partition functions for the indicated pairs. Equation (4.2.14) reduces to

$$r^2 e^{2\epsilon/kT} = r^2 - 2r + 4xy \quad (4.2.15)$$

The proper solution for r from this equation is that given in Eq. (4.2.12). It is in this connection that the first approximation is often referred to as the quasi-chemical approximation (QCA).

QCA represents an improvement over the zeroth approximation because in it the correlation of the distribution of A and B atoms, called the correlation state of the alloy depends on the pair energies. The level of improvement can be appreciated by comparison with the exact analytical solutions in two dimensions. The binary alloy statistical problem with pair interactions is equivalent to the magnetic Ising model. For $\epsilon > 0$, the alloy problem can be directly translated into the ferromagnetic Ising model. The only difference is that while the alloy concentration χ is specified, x in the magnetic problem is further varied to minimize ΔF . Thus above T_c and with pair interactions the minimum of ΔF always occurs at $x = 1/2$, which in the magnetic case means no net spin or magnetic moment. However, for temperatures below T_c , ΔF has two minima occurring at x values x_1 and x_2 corresponding to the miscibility gaps indicated in Fig. 4.1.1. For magnetism, this situation corresponds to having a net magnetic moment. Thus the alloy spinodal critical temperature T_c is equivalent to the magnetic phase transition critical temperature.

The value of T_c in the pair model can be written as

$$T_c = \frac{\Omega}{\lambda R}, \quad (4.2.16)$$

where Ω is the interaction parameter $\Omega = N_0 z \epsilon$ defined in Eq. (4.2.7), and λ is a numerical factor that depends on the model used. For the zeroth approximation we already found $\lambda = 2$, [see Eq. (4.2.8)]. The zeroth approximation is equivalent to the Bragg-Williams (1954) approximation to the Ising model. For QCA, the λ value is given by

$$\lambda = z \ln \left(\frac{z}{z-2} \right). \quad (4.2.17)$$

This result can be obtained by starting from the expression for ΔF in Eq. (4.2.11) and r in Eq. (4.2.12), and then taking the second derivative $\partial^2 \Delta F / \partial x^2$ and setting it equal to zero at $x = 1/2$. In doing so, the conditions $\partial \Delta F / \partial r = 0$ and $\partial r / \partial x = 0$ at $x = 1/2$ are used. An explicit derivation of a more accurate

expression is given in Sec. 4.7. Note that QCA for alloys is equivalent to the Bethe-Peierls approximation (Bethe, 1935; Huang, 1987) to the Ising model.

Table 4.2.1 compares the λ values that enter the T_c formula, Eq. (4.2.17), for a square lattice from different approximations against the exact solution (Onsager, 1944) and for a fcc lattice against the best numerical value available (de Fontaine, 1979). The results obtained from the cluster variation method (CVM) to be discussed in Sec. 4.10 are also listed for later use. Progressively more sophisticated statistical theories result in

Table 4.2.1

The values of λ for the transition temperature $T_c = \Omega/(\lambda R)$ from different calculations. CVM indicates the results from the cluster variation method to be discussed in Sec. 4.10.

Alloy, Ising Model	Zeroth Approx. Bragg-Williams	QCA, Bethe-Peierls	CVM	Exact or Best Result
Square Lattice	2	2.773	3.332 (square)	3.5255
fcc	2	2.188	2.438	2.4501

After de Fontaine (1974), Table V.

larger λ values, and hence smaller critical temperatures. Although QCA predicts only 79% of the exact λ value in the square lattice case, it gives 89% of the best computer simulated value for the fcc lattice. Therefore not surprisingly, QCA has been used with reasonable success in semi-empirical phase diagram calculations for semiconductor compounds and alloys, as will be discussed in the next few sections.

4.3 Phase Diagram: Common Tangent Line and Activity Coefficient

So far our calculation of free energies has been confined to solid solutions. To find the phase diagrams one also needs to know the free energies of liquids (or melts). Very often, liquids are treated as dense lattice gases, as solids with a high concentration of vacancies, or simply as a disordered system with some effective coordination number. The latter treatment has been used with some success in the calculation of phase diagrams of compound semiconductors and alloys. This and the next sections introduce the basic ideas governing phase diagrams including (1) the liquidus curve between a binary liquid $A_{1-x}B_x$ and a stoichiometric compound AB, e.g., between a $Ga_{1-x}As_x$ melt and stoichiometric solid GaAs; (2) the liquidus-solidus curves describing the equilibrium between a ternary liquid and a solid pseudobinary alloy; and (3) the general ternary phase diagrams. The basic ideas and most results are in the papers of Vieland (1963), Stringfellow and Green (1969), Kikuchi (1981 and 1982) and Breberick et al. (1983), and in a text by Casey and Panish (1978).

As mentioned earlier the equilibrium state at a given P and T is the state with minimum Gibbs free energy. To construct a phase diagram, we first calculate the free energy curves as a function of alloy concentrations for all the relevant phases, and then find the combination of phases that minimizes the free energy. For example, consider the liquidus-solidus curves between a pseudobinary liquid and the solid solutions $A_{1-x}B_x$ C schematically shown in Fig. 4.3.1a. The corresponding free energy curves for both phases as a function of x for a given temperature T_1 are drawn in Fig. 4.3.1b. Pay attention to the common tangent line that touches the two free energy curves at x_1 and x_2 . This picture illustrates the cause of the equilibrium state behavior as a function of the average alloy concentration x at this temperature. For x less than x_1 the liquid phase is the stable phase, and for $x > x_2$ the solid solution is the stable phase. For an average concentration lying between x_1 and x_2 the stable state is a mixture of a liquid phase with concentration x_1 and a solid phase of

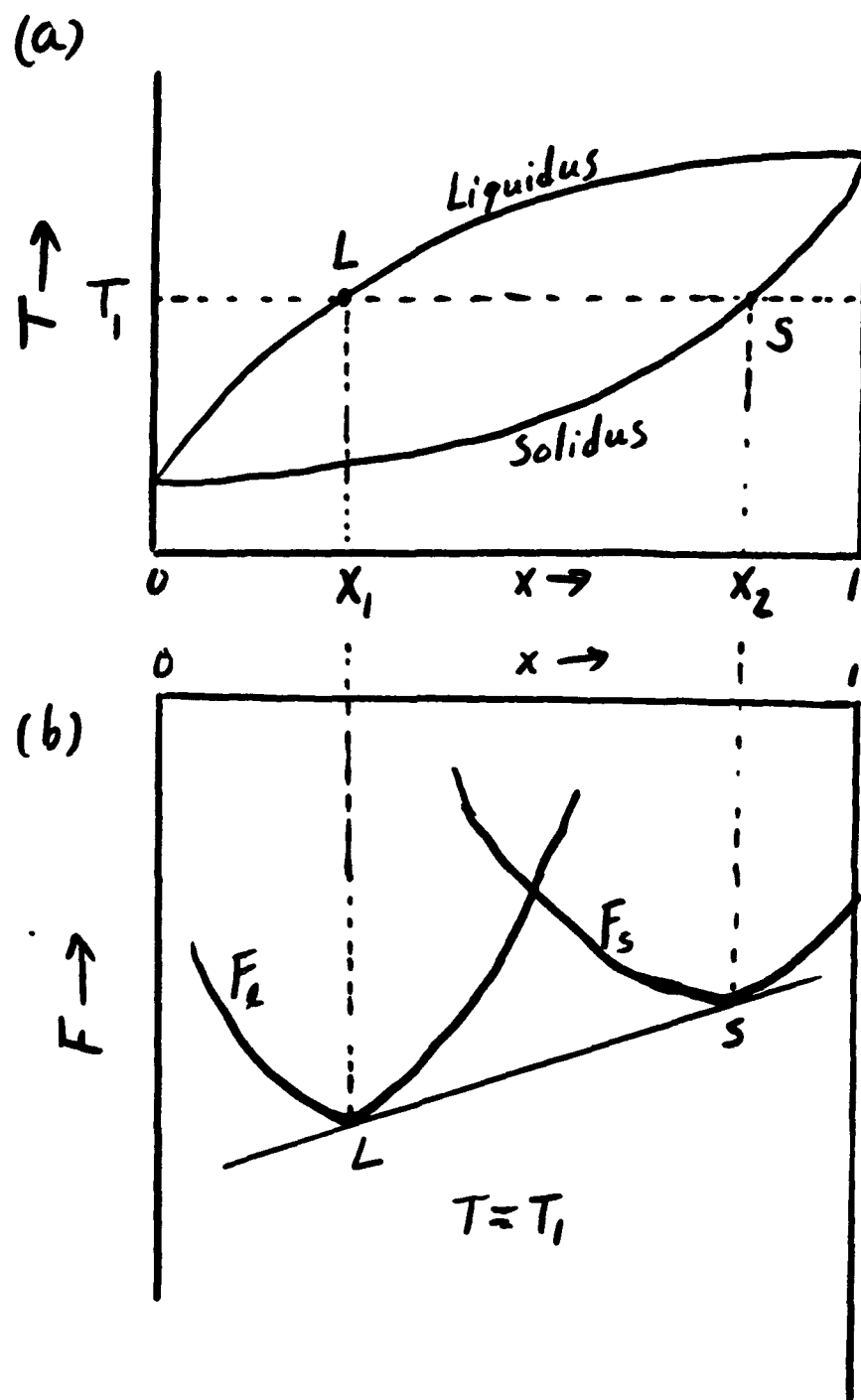


Figure 4.3.1 (a) Schematic liquidus and solidus curves, and (b) the corresponding free energies of the liquid F_L and solid F_S at T_1 as a function of alloy composition x .

concentration x_2 with proportions governed by the level rule. The two free energy curves at this temperature T_1 give two phase boundary points (denoted by L and S) for the liquidus-solidus curves shown in Fig. 4.3.1a. The full liquidus-solidus curves are generated from a set of such free energy curves derived by continuously varying the temperature. This procedure for minimizing the free energy using the common tangent line is equivalent to matching the chemical potentials, as will be demonstrated below.

The Gibbs free energy G , or in practice the Helmholtz free energy F , for a given phase can be written as a sum of the products of the number of particles N_i and the chemical potentials μ_i of all the atom and molecular species i in that phase,

$$F = G = \sum_i N_i \mu_i \quad (4.3.1)$$

Conversely the chemical potential can be calculated from

$$\mu_i = (\partial F / \partial N_i)_{(N_j, i \neq j)} \quad (4.3.2)$$

Now consider the equilibrium between two phases, denoted by α and β , of a binary alloy $A_{1-x}B_x$ at a given temperature T , such as depicted in Fig. 4.3.1b. Let the common tangent line touch the two free energy curves F_α and F_β at x_α and x_β respectively. Denote the chemical potential of an A atom in the α phase $\mu_A^\alpha(x_\alpha, T)$ as μ_A^α , and assign similar meanings to μ_B^α , μ_A^β , and μ_B^β . Algebraically, the common tangent line requires

$$\partial F_\alpha(x_\alpha, T) / \partial x = \partial F_\beta(x_\beta, T) / \partial x \quad (4.3.3)$$

and

$$F_\beta(x_\beta, T) = F_\alpha(x_\alpha, T) + (x_\beta - x_\alpha) \partial F_\alpha(x_\alpha, T) / \partial x \quad (4.3.4)$$

In terms of the chemical potentials, we have

$$F_\alpha(x_\alpha, T) = N [x_\alpha \mu_B^\alpha + (1 - x_\alpha) \mu_A^\alpha], \quad (4.3.5)$$

$$F_\beta(x_\beta, T) = N [x_\beta \mu_B^\beta + (1 - x_\beta) \mu_A^\beta], \quad (4.3.6)$$

$$\partial F_{\alpha} (x_{\alpha}, T) / \partial x = N(\mu_B^{\alpha} - \mu_A^{\alpha}), \quad (4.3.7)$$

$$\partial F_{\beta} (x_{\beta}, T) / \partial x = N(\mu_B^{\beta} - \mu_A^{\beta}), \quad (4.3.8)$$

Using these expressions in Eqs. (4.3.3) and (4.3.4) leads to the usual conditions for chemical equilibrium, $\mu_A^{\alpha} = \mu_A^{\beta}$ and $\mu_B^{\alpha} = \mu_B^{\beta}$.

Very often the activity coefficients γ rather than the chemical potentials μ are used in phase diagram evaluations. The activity coefficient γ_i is related to the chemical potentials by

$$\mu_i - \mu_i^0 = kT \ln(x_i \gamma_i), \quad (4.3.9)$$

where μ_i^0 is the chemical potential for the pure phase of the i^{th} species. For an ideal solution, i.e. a random mixture which is always found in the limit $T \rightarrow \infty$, all the activity coefficients become unity, $\gamma_i = 1$. In the zeroth approximation for a binary $A_{1-x}B_x$ (or a pseudobinary $A_{1-x}B_xC$), the activity coefficients can be shown to be given by

$$\gamma_A = e^{x^2 \Omega / (RT)}, \quad (4.3.10)$$

$$\gamma_B = e^{(1-x)^2 \Omega / (RT)}. \quad (4.3.11)$$

In QCA, using Eq. (4.2.11) and (4.2.12), one finds

$$\gamma_A = \left(\frac{1 - 2x + f}{(1-x)(1+f)} \right)^{1/2} \quad (4.3.12)$$

$$\gamma_B = \left(\frac{2x - 1 + f}{x(1+f)} \right)^{1/2} \quad (4.3.13)$$

where f is related to the excess energy ϵ by

$$f = [1 + 4xy (e^{2\epsilon\beta} - 1)^{1/2}] \quad (4.3.14)$$

with $\beta = 1/(kT)$. These results will be used in the next two sections.

4.4 Vieland's Method and a Binary Liquidus

The Vieland (1963) method, which establishes the relation between chemical potentials of a stoichiometric solid AB compound and those of the super-cooled liquid solution $A_{0.5}B_{0.5}$ through the entropy of fusion and heat capacity differences between the two phases, remains the principal way of calculating phase diagrams between semiconductor solids and liquids.

Consider the equilibrium between a liquid solution $A_{1-x}B_x$ and a solid compound AB. A schematic liquidus curve is shown in Fig. 4.4.1a while the corresponding free energy diagram at a given temperature T is given in Fig. 4.4.1b. In equilibrium, the chemical potential μ_{AB} per AB pair (unit cell) in the AB compound must be equal to the sum of the chemical potentials in the liquid solution

$$\mu_{AB} = \mu_A^0(x_0) + \mu_B^0(x_0) , \quad (4.4.1)$$

where x_0 is the point at which the common tangent line touches the liquid free energy curve shown in Fig. 4.4.1b. The free energy of the solid compound with N unit cells at this temperature is of course $F_s = N\mu_{AB}$. F_s can be related to the free energy of a super-cooled 50 - 50 liquid solution by following a sequence of quasi-equilibrium steps. First, heat one mole of an AB crystal from T to the melting temperature T_m , i.e., going from the point marked S to M in Fig. 4.4.1a. The change in the free energy in this step is

$$F_s(T_m) - F_s(T) = - \int_T^{T_m} S_s(T') dT' , \quad (4.4.2)$$

The subscript s denotes the solid phase. The entropy S is related to the heat capacity C_s by

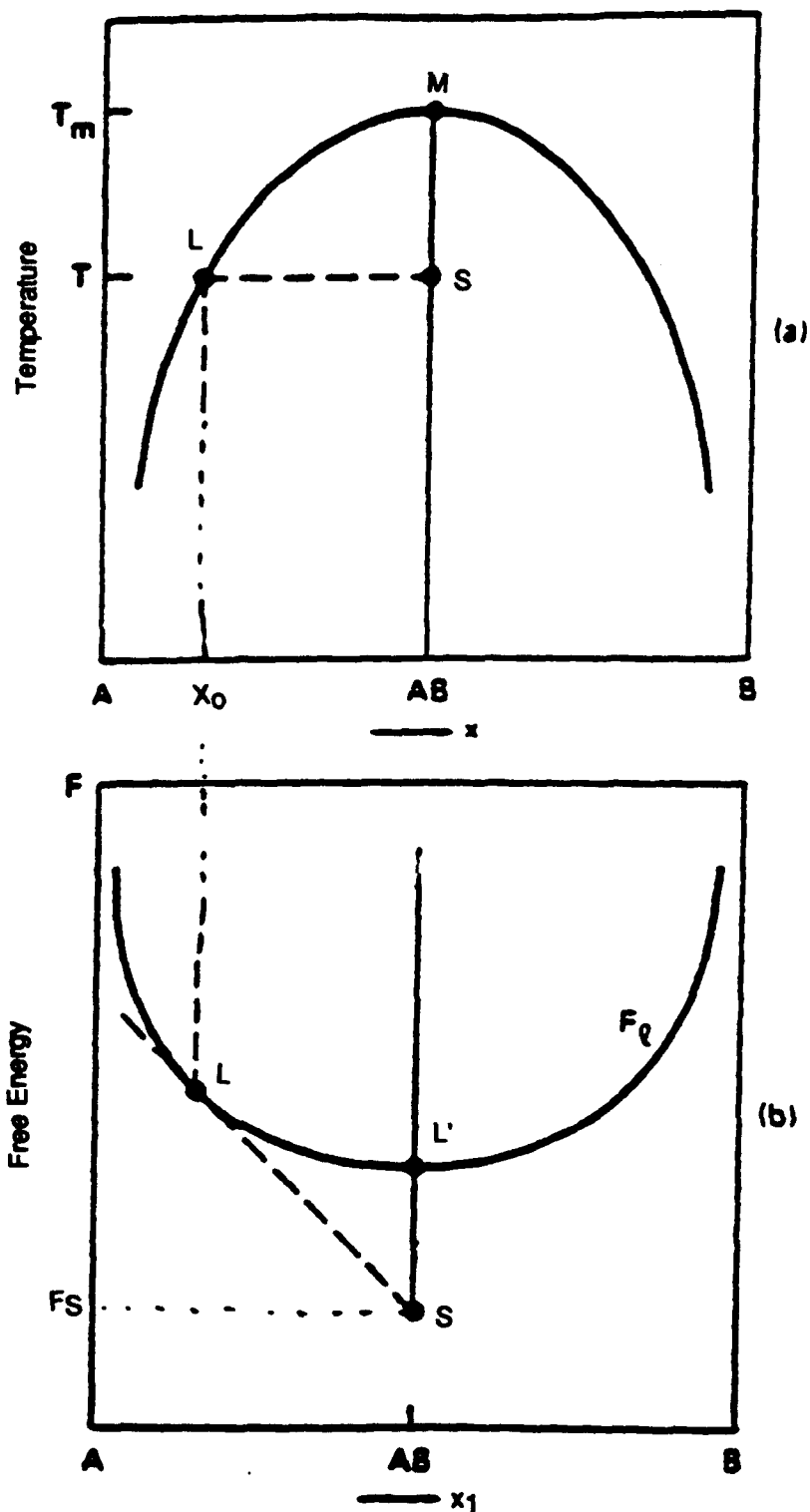


Figure 4.4.1 (a) Schematic liquidus curve for a compound semiconductor, (b) free energy curve F_L as a function of composition x of a binary liquid at a given temperature T and the corresponding value F_S for the solid compound at the same temperature. Points L and S coexist.

$$S_s(T) = S_s(T_m) - \int_T^{T_m} \frac{C_s(T')}{T'} dT' \quad (4.4.3)$$

Combining the two equations above yields,

$$F_s(T_m) - F_s(T) = -S_s(T_m)(T_m - T) + \int_T^{T_m} \int_T^{T_m} C_s(T'') \frac{1}{T''} dT'' dT' \quad (4.4.4)$$

The second step is to melt the crystal at T_m which causes no change in the free energy. This situation corresponds to Figure 4.4.1b with the liquid free energy curve lowered to pass through the solid free energy minimum labeled S.

Finally the liquid is cooled back from T_m to T (a super-cooled liquid), which changes the free energy by

$$F_l(T) - F_l(T_m) = -S_l(T_m)(T - T_m) - \int_T^{T_m} \int_T^{T_m} C_l(T'') \frac{1}{T''} dT'' dT' \quad (4.4.5)$$

The subscript l indicates the liquid. The total change is

$$\begin{aligned} F_l(T) - F_s(T) &= [S_l(T_m) - S_s(T_m)](T_m - T) \\ &+ \int_T^{T_m} \int_T^{T_m} [C_l(T'') - C_s(T'')] \frac{1}{T''} dT'' dT' \end{aligned} \quad (4.4.6)$$

In terms of chemical potentials, Eq. (4.4.6) can be written for one mole of AB compound as,

$$\begin{aligned} N_0[(\mu_A(0.5) + \mu_B(0.5) - \mu_{AB})] &= N_0 \Delta\mu_{ls}^{AB} \\ &= \Delta S_m(T_m - T) - \Delta C[T_m - T - T \ln(T_m/T)], \end{aligned} \quad (4.4.7)$$

where $\Delta C = C_l(T) - C_s(T)$ is assumed to be independent of T , and the entropy of fusion is defined as $\Delta S_m = S_l(T_m) - S_s(T_m)$ for one mole of AB compound. Using Eq. (4.4.1) for μ_{AB} and the relation between the chemical potentials and activity coefficients in Eq. (4.3.9), Eq. (4.4.7) can now be written as

$$\ln \left[\frac{1}{4x_0(1-x_0)} \right] + \ln \left[\frac{\gamma_A(0.5) \gamma_B(0.5)}{\gamma_A(x_0) \gamma_B(x_0)} \right] \quad (4.4.8)$$

$$= \frac{\Delta S_m}{RT} (T_m - T) - \frac{\Delta C}{RT} \left[T_m - T - T \ln \left(\frac{T_m}{T} \right) \right] .$$

This is Vieland's (1963) formula. It is a simple transcendental equation for T as a function of X_0 - the liquidus curve. When the liquid solution is in equilibrium with the pure solid B phase, the liquidus is governed by a similar equation

$$\ln \left[\frac{1}{x_0 \gamma_B(x_0)} \right] \frac{\Delta S_m^B}{RT} (T_m^B - T) - \frac{\Delta C^B}{RT} \left[T_m^B - T - T \ln \left(\frac{T_m^B}{T} \right) \right] . \quad (4.4.9)$$

To actually calculate the liquidus curve using Eqs. (4.4.8) and (4.4.9), one needs to know the activity coefficients γ_A and γ_B , which of course, depend on the statistical model and the energetics. If QCA is used, these activity coefficients are given explicitly by Eqs. (4.3.12) and (4.3.13). Then the remaining quantities needed are the mixing energy parameter $\Omega = N_0 z \epsilon$ for the binary liquid, the fusion entropy ΔS_m , and the heat capacity difference ΔC . In most applications ΔC has been taken to be zero, which is a good approximation at high temperature. This approach works reasonably well for the III-V binary compounds. Examples are shown in Fig. 4.4.2, where calculated liquidus curves (Stringfellow and Green, 1969; Kikuchi, 1981) for GaAs, InAs and InSb are compared with experiment. The parameters used in the calculations are listed in Table 4.4.1. Note that the entropy unit eu used for ΔS_m is eu = J/(K-mole). The symmetrical liquidus curves

fig 4.4.2

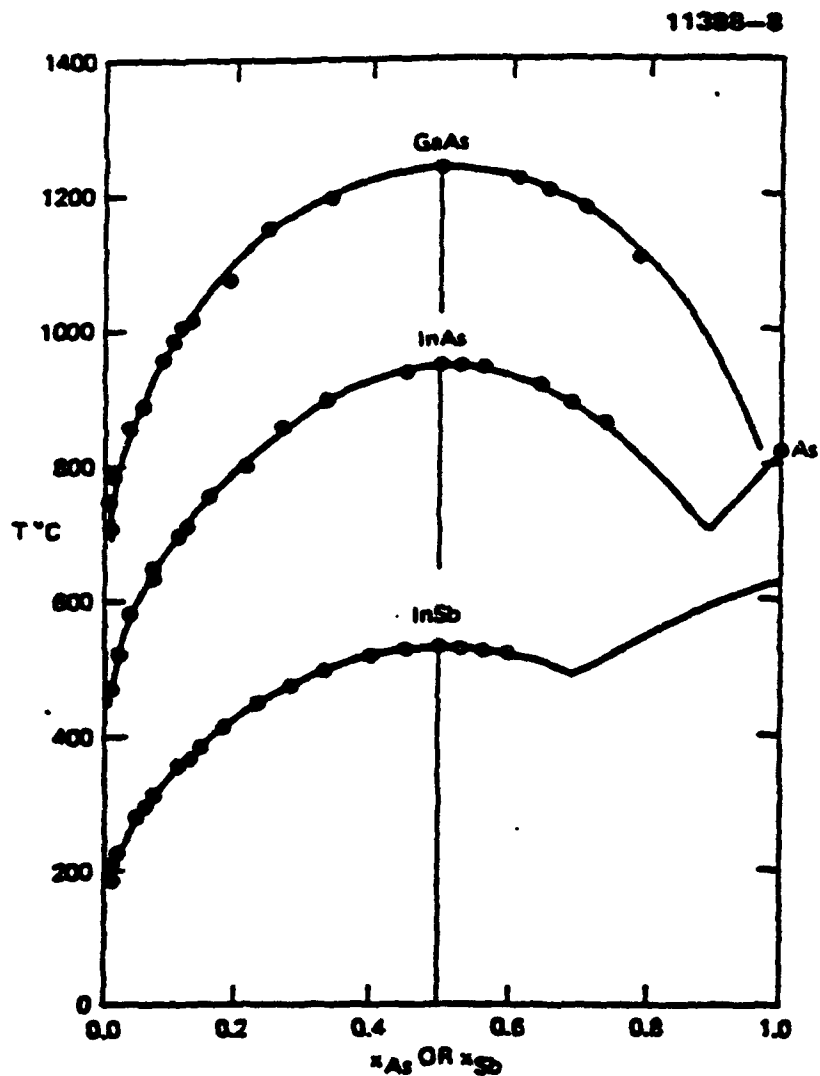


Figure 4.4.2 Liquidus curves of III-V semiconductors. Solid curves are theory and circles are experiments. After Kikuchi (1981).

indicate that the QCA model, with coordination number $z = 4$, is a reasonable representation of these binary melts. This approach, however, does not work well for the II-VI systems. As shown in Fig. 4.4.3, the skewed and sharp nature of the experimental curves indicates that these liquid solutions are more complex than the simple picture represented by QCA for the binary solutions. A simple extension of QCA to include the molecule species AB in addition to A and B in the liquid solutions is found to work well, as indicated by the comparison with experiment in Fig. 4.4.3. This extended QCA model is called the associated solution model. Kikuchi (1982) found that repulsive interactions between the molecular and atomic species in the solution are responsible for the skewed nature of the liquidus curves.

Table 4.4.1

Parameters used in the calculation of liquidus curves in Fig. 4.4.2 and phase diagrams in Fig. 4.5.4.

Melting Temperature T_m (K)		Entropy of fusion $\Delta S_m(\text{eu})$		Mixing enthalpy parameter Ω (cal/mole)	
GaAs	1511	GaAs	14.7	Ga-As	-4380
InAs	1215	InAs	14.7	In-As	-6070
InSb	803	InSb	13.3	In-Sb	-3980
As	1090	As	3.67	In-Ga	1066
Sb	903	Sb	5.27	As-Sb	610
				GaAs-InAs	2800
				InAs-InSb	2900

After Stringfellow and Green (1969)

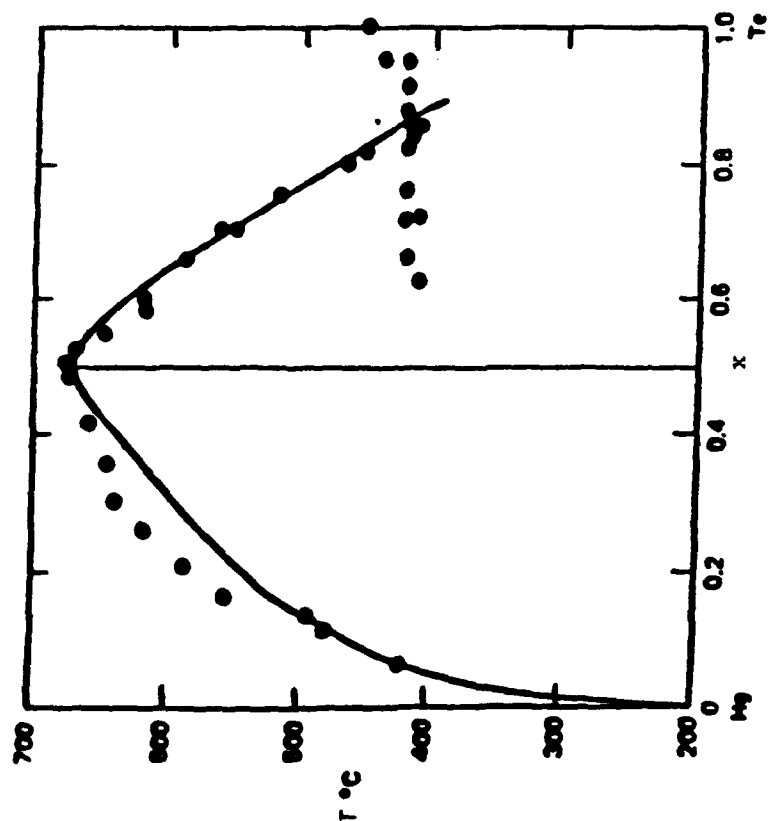
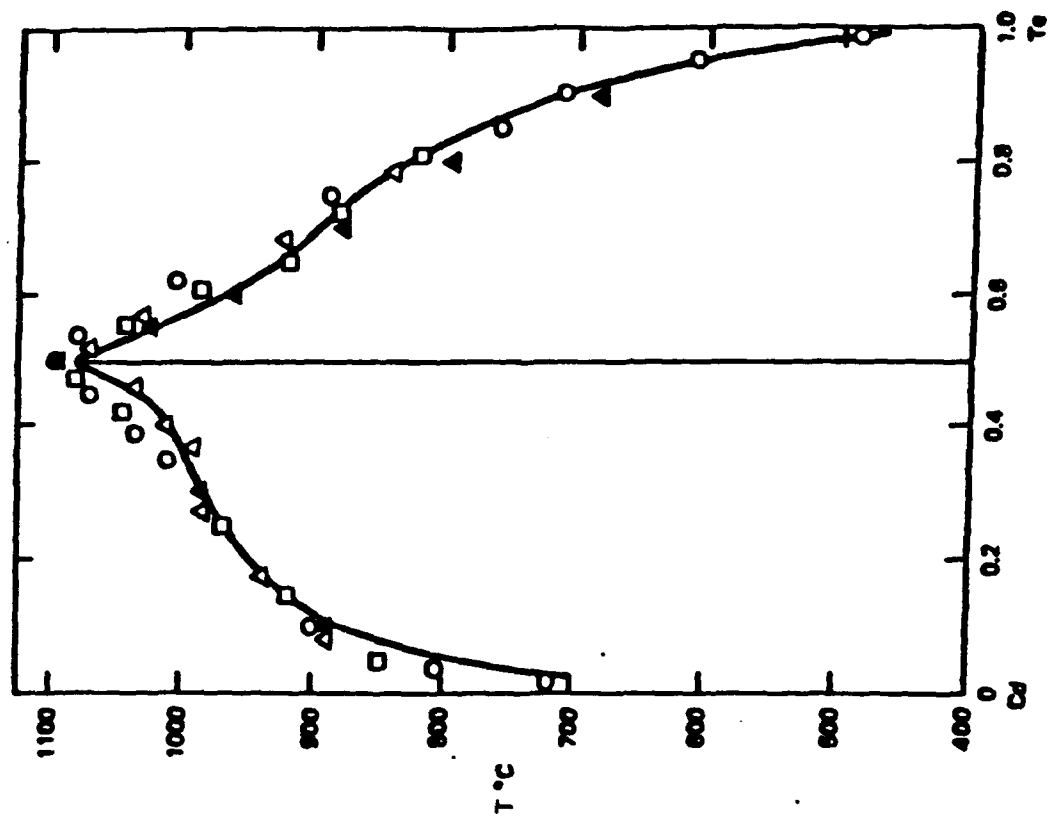


Figure 4.4.3 Binary liquidus. The curve is the theory and the points are experiments for (a) CdTe and (b) Hg-Te. After Kikuchi (1982b).

4.5 Ternary Phase Diagram

The methods of Sec. 4.4 are next extended to the phase diagrams for ternary systems. The phase boundary curves are the solidus-liquidus curves that describe the equilibrium between the pseudo-binary solid solutions

$A_{1-x}B_xC$ and the liquid ternary solution denoted by $A_aB_bC_c$, where a , b , and c are fractional concentrations satisfying normalization $a+b+c = 1$. A schematic three-dimensional phase diagram is shown in Fig. 4.5.1; it is a plot of the temperature T versus concentrations. The liquidus is the upper surface and the solid pseudobinary phase is the plane lying between the two lines labeled AC and BC . A given set of a, b , and c values corresponds to a point D inside the base triangle ABC , as shown in Fig. 4.5.2. The values of a , b , and c are related to the geometric line ratios by $a = DA'/AA'$, $b = DB'/BB'$ and $c = DC'/CC'$. The solidus-liquidus boundaries form a family of curves relating the solid concentration x and the temperature T to the two independent liquid concentrations a and b (note $c = 1 - a - b$). These relations can be obtained by matching the chemical potentials μ described below.

Let us start with μ_{AC} and μ_{BC} for the AC and BC components in the solid solution $A_{1-x}B_xC$. They are related to those in the pure AC and BC compounds through the activity coefficients γ by

$$\mu_{AC}(x,T) = \mu_{AC}^0(T) + kT \ln[(1-x)\gamma_{AC}] , \quad (4.5.1)$$

and

$$\mu_{BC}(x,T) = \mu_{BC}^0(T) + kT \ln(x\gamma_{BC}) , \quad (4.5.2)$$

where the superscript 0 designates pure compounds. Now Vieland's relation in Eq. (4.4.7) can be used to relate the chemical potentials μ^0 of the pure solid compounds to those of the super-cooled 50-50 binary liquid mixtures, denoted μ^{sc} . For example, μ_{AC}^0 is given by

$$\mu_{AC}^0 = \mu_A^{sc} + \mu_C^{sc} - \Delta\mu_{AC}^{sc} . \quad (4.5.3)$$

Fig. 4.5.1

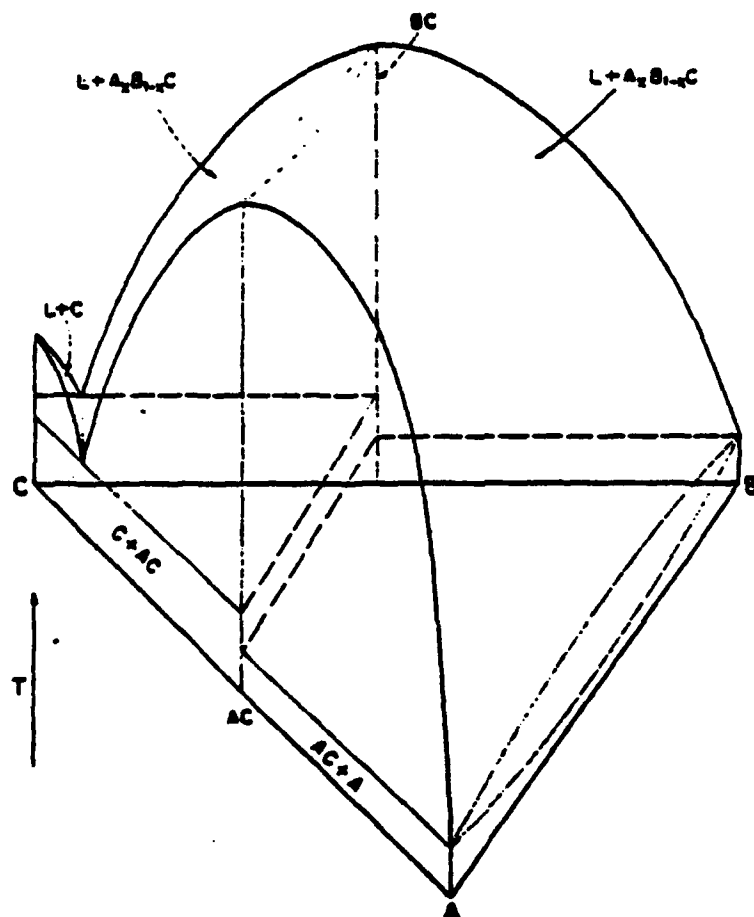


Figure 4.5.1 Schematic A-B-C ternary phase diagram. After Stringfellow and Green (1969).

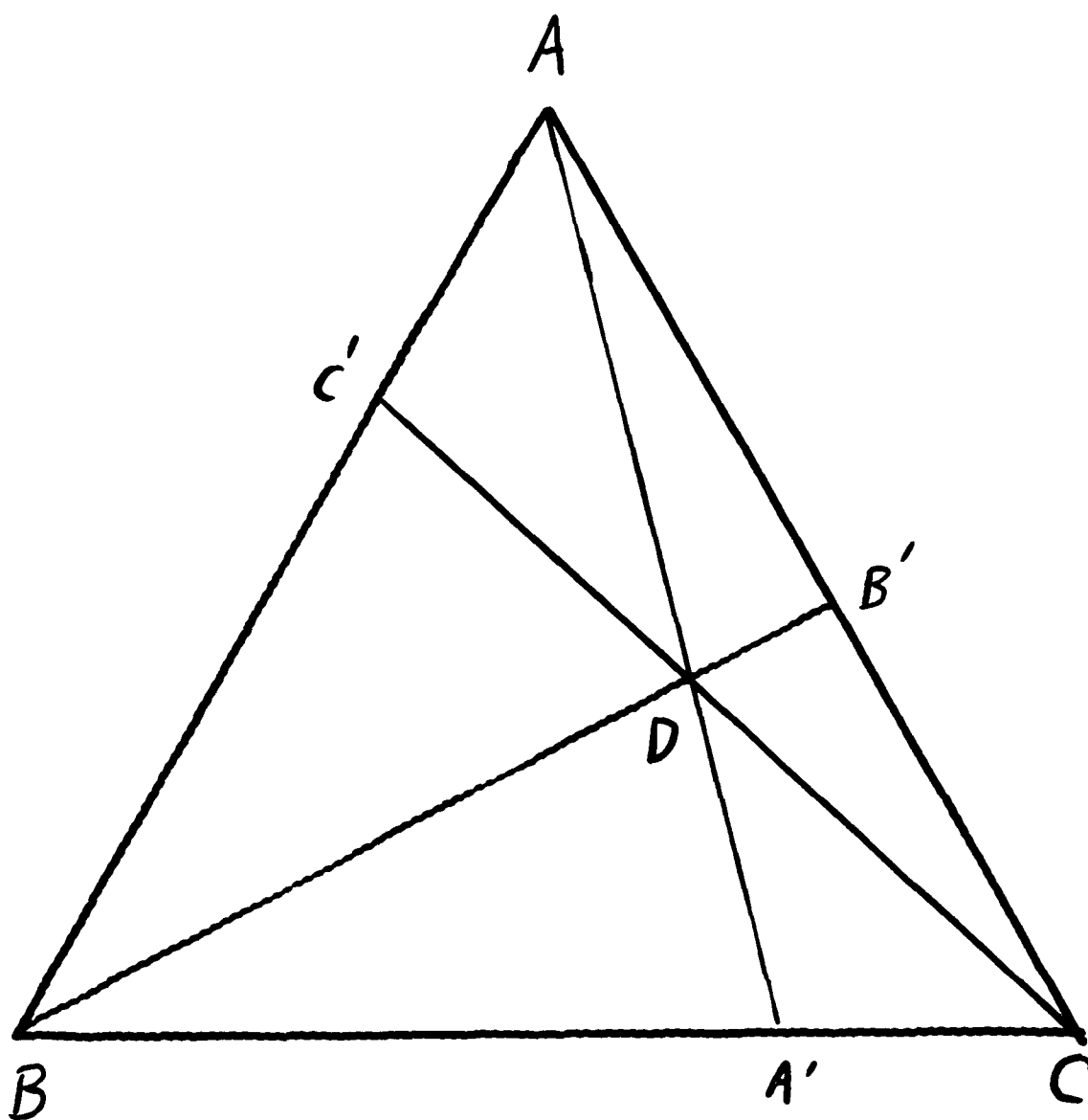


Figure 4.5.2 Any point D inside the triangle ABC defines a unique set of fraction concentrations in ternary alloy $A_aB_bC_c$ by $a = DA'/AA'$, $b = DB'/BB'$, and $c = DC'/CC'$.

Note that $\Delta\mu$ is expressed in terms of a change in the heat capacity ΔC and entropy of fusion ΔS_m between the 50-50 binary liquid and the solid compound AC given by the right side of Eq. (4.4.7). Since the ternary liquid is in equilibrium with the solid pseudo-binary, μ_{AC} and μ_{BC} in the latter are equal to the sum of the chemical potentials for the atomic species in the former by

$$\mu_{AC}(x,T) = \mu_A^l(a,b,T) + \mu_C^l(a,b,T) . \quad (4.5.4)$$

A similar expression holds for μ_{BC} . The μ^l for the atomic species in the ternary liquid can be further expressed in terms of those of the pure elementary liquids and activity coefficients by

$$\mu_A^l = \mu_A^0 + kT \ln(a\gamma_A) . \quad (4.5.5)$$

Combining the above three equations with Eq. (4.5.1), one obtains the equilibrium condition

$$\ln[(1-x)\gamma_{AC}] = \ln\left(\frac{a\gamma_A\gamma_C}{\gamma_A^{sc}\gamma_C^{sc}}\right) + \frac{\Delta\mu_{ls}^{AC}}{kT} . \quad (4.5.6)$$

Similarly, Eq. (4.5.2) leads to

$$\ln(x\gamma_{BC}) = \ln\left(\frac{b\gamma_B\gamma_C}{\gamma_B^{sc}\gamma_C^{sc}}\right) + \frac{\Delta\mu_{ls}^{BC}}{kT} . \quad (4.5.7)$$

To calculate the phase diagram from Eqs. (4.5.6) and (4.5.7), one needs to know the x and T dependence of the activity coefficients γ_{AC} and γ_{BC} for the pseudobinary solid solutions, the T dependence of the activity coefficients γ^{sc} of the super-cooled 50-50 binary liquids, and the T and concentration (a, b , and c) dependences of γ_A , γ_B , and γ_C in the ternary liquid solutions $A_aB_bC_c$. In addition ΔC and ΔS_m are needed for the calculation of the $\Delta\mu$. If QCA is used, then the activity coefficients for the binary liquid and pseudobinary solid solutions take the forms of Eqs. (4.3.12) and (4.3.13). For the ternary liquid solution, QCA has to be extended to three components. There are now six different kinds of pairs to be considered, namely, AA,

BB, CC, AB, BC and AC. We label these pairs by j , where j runs from 1 to 6, and denote the fractions of each pair as x_j . If independent pair energies $\epsilon_{\alpha\beta}$ are assumed, there are only three nonvanishing energy parameters, Δ_{AB} , Δ_{AC} , and Δ_{BC} , while $\Delta_{AA} = \Delta_{BB} = \Delta_{CC} = 0$, where we have defined $\Delta_{\alpha\beta} = \epsilon_{\alpha\beta} - (\epsilon_{\alpha\alpha} + \epsilon_{\beta\beta})/2$. Then it is straightforward to write the mixing energy as

$$\Delta E = \sum_{j=1}^6 M x_j \Delta_j, \quad (4.5.8)$$

where $M = Nz/2$ with z being an effective coordination number, taken to be 4. The mixing entropy is $\Delta S = k \ln \Phi$, with Φ given by

$$\Phi = \frac{N!}{N_A! N_B! N_C!} \frac{M!}{\prod_{j=1}^6 M_j!} \prod_{j=1}^6 [(x_j^0)^{M_j}], \quad (4.5.9)$$

where $M_j = x_j M$ and x_j^0 are the *a priori* probabilities, e.g., $x_{AA}^0 = a^2$ and $x_{AB}^0 = 2ab$ etc. This Φ is a simple extension of Eq. (4.2.9). The mixing free energy then reads

$$\Delta F = \sum_{j=1}^6 M [x_j \Delta_j + k T X_j \ln(x_j / x_j^0)] + N k T (a \ln a + b \ln b + c \ln c). \quad (4.5.10)$$

Note that not all the x_j are independent. They are constrained to have the right numbers of A B, and C atoms for a given set of fractional concentrations, a , b , and c :

$$2x_{AA} + x_{AB} + x_{AC} = 2a, \quad (4.5.11)$$

$$2x_{BB} + x_{AB} + x_{BC} = 2b, \quad (4.5.12)$$

$$2x_{CC} + x_{CB} + x_{AC} = 2c, \quad (4.5.13)$$

Minimization of ΔF with respect to the x_j under the imposed constraints leads to the following three coupled equations:

$$(2a - x_{AB} - x_{AC})(2b - x_{AB} - x_{BC}) = x_{AB}^2 \exp[\Omega_{AB} / (2RT)], \quad (4.5.14)$$

$$(2a-x_{AB} - x_{AC}) (2c-x_{AC} - x_{BC}) = x_{AC}^2 \exp [\Omega_{AC} / (2RT)] , \quad (4.5.15)$$

$$(2b-x_{AB} - x_{BC}) (2c-x_{AC} - x_{BC}) = x_{BC}^2 \exp [\Omega_{BC} / (2RT)] , \quad (4.5.16)$$

where the mixing energy parameters are defined as $\Omega_{\alpha\beta} = N_0 z \Delta_{\alpha\beta}$, with z taken to be four. After Eqs. (4.5.11) through (4.5.16) are solved for x_j , they are used in Eq. (4.5.10) to obtain the functional dependence of ΔF on the concentrations a , b , and c and on the temperature T . Then the activity coefficients can be calculated. For example, γ_A is calculated from

$$\begin{aligned} kT \ln(a\gamma_A) &= (\partial \Delta F / \partial N_A)_{N_B, N_C} \\ &= (\partial \Delta F / \partial a) (\partial a / \partial N_A) + (\partial \Delta F / \partial b) (\partial b / \partial N_A) . \end{aligned} \quad (4.5.17)$$

In this equation, a and b have been chosen as independent variables with $c = 1 - a - b$.

Thus the parameters needed to complete the specification of a ternary phase diagram in QCA are ΔC and ΔS_m for both AC and BC compounds, the mixing energy parameters Ω_{AB} , Ω_{AC} , and Ω_{BC} for the ternary liquid and $\Omega_{AB(C)}$ for the pseudobinary solid solution. It is a good approximation to set $\Delta C = 0$. There are four independent variables, say, a , b , x , and T , and two equations, Eqs. (4.5.6) and (4.5.7) with which to work. One can fix T and solve for a (and hence $c = 1 - a - b$) and x as functions of b . Plots of the contour of a , b , and c for the same T in the Gibbs triangle ABC, is an isotherm. If the solidus concentration x is fixed, the contour of a , b , and c in the Gibbs phase plot is the iso-solidus-concentration line. Figure 4.5.3 shows both sets of these plots for the In-Ga-As system reported by Stringfellow and Green (1969) using the parameters given in Table 4.4.1. We can fix $c = 0.5$ in the ternary and solve for x and T as a function of a , then we obtain the liquidus and the solidus curves along the $c = 1/2$ line, as shown in Fig. 4.5.4, also by Stringfellow and Green and further verified by Kikuchi (1981). The agreement between the theory and experiment is remarkably good.

III-V TERNARY PHASE DIAGRAMS

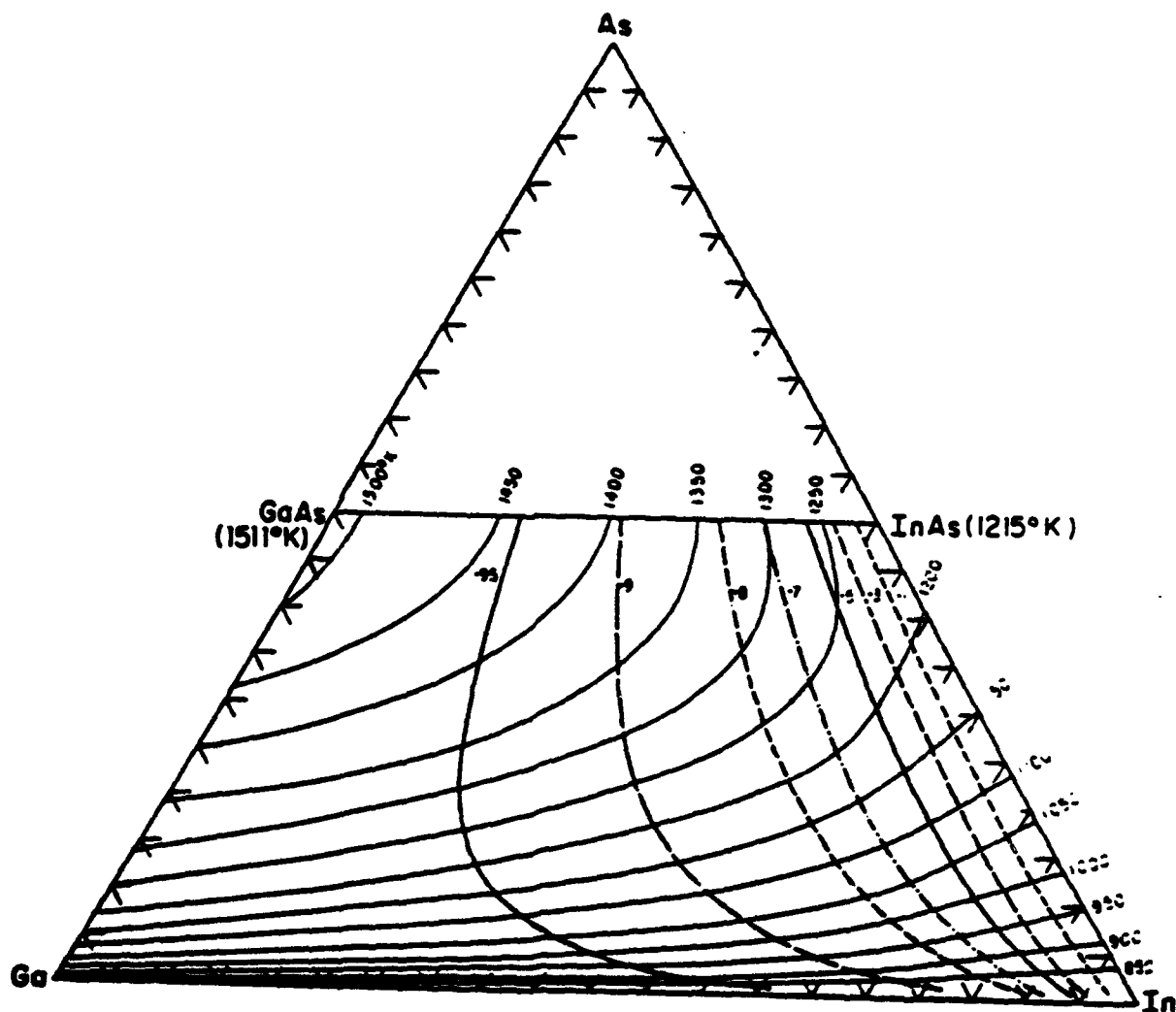


Figure 4.5.3 Isotherms and iso-solidus concentration lines for phase equilibrium between pseudobinaries $\text{In}_{1-x}\text{Ga}_x\text{As}$ and the corresponding ternary liquid. After Stringfellow and Green (1969).

4.5.4

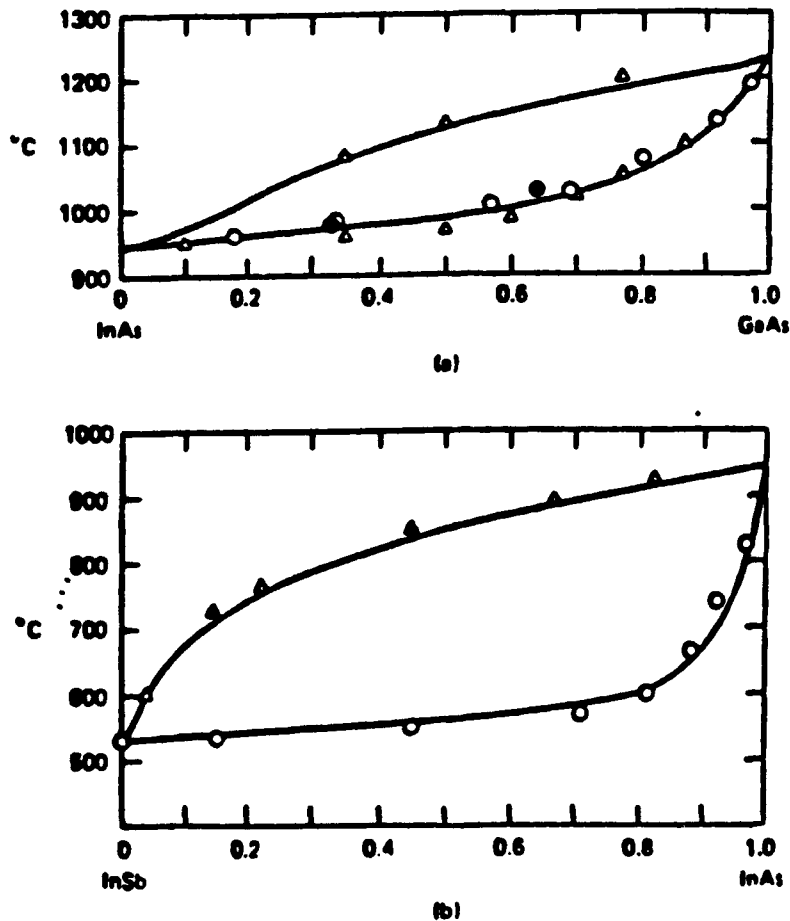


Figure 4.5.4 Liquidus and Solidus curves for $\text{In}_{1-x}\text{Ga}_x\text{As}$ and $\text{InSb}_{1-x}\text{As}_x$. Solid curves are calculations and the circles are the experiments. After Kikuchi (1981).

4.6 Phase-Diagram Data and Simple Mixing Enthalpy Models

Before moving into more rigorous statistical theory and more systematic energetic calculations, it is useful to summarize the semiempirical phase diagram data and simple statistical and energetic models that have been used in practical calculations.

From a survey of phase diagram calculations of ternary and quaternary semiconductors, Stringfellow (1974) preferred the following forms for mixing free energies. For the III-V liquid solutions, the preferred form (per mole) is

$$\Delta F_l = x(1-x) (\xi_1 - \xi_2 T) + RT (x \ln x + y \ln y) \quad (4.6.1)$$

For the III-III and V-V liquids and pseudobinary solids, the strict regular solution model is his preference,

$$\Delta F = x(1-x)\Omega + RT (x \ln x + y \ln y) \quad (4.6.2)$$

Table 4.6.1 lists the values of $\Omega_l = \xi_1 - \xi_2 T$ and other phase diagram data including the entropy of fusion ΔS_m , the melting temperatures T_m for the III-V binary systems, and the mixing energy parameters for the III-III and V-V interactions in the liquid solutions. In Table 4.6.2, the experimental values of Ω based on Eq. (4.6.2) are compared with the mixing enthalpies calculated from several theoretical models to be discussed presently. One has to be cautious in comparing the experimental Ω with the calculated values, because the former may include additional contributions from entropy terms not contained in the second term of Eq. (4.6.2).

One simple model for estimating the mixing energy is the so-called delta lattice parameter model (DLP) suggested by Stringfellow (1974). He observed that the experimental Ω values for the III-V pseudobinary alloys $A_{1-x}B_xC$ correlated strongly with the difference in the lattice constant $\Delta a = a_{BC} - a_{AC}$, as shown in Fig. 4.6.1 in a log-log plot generated by

Table 4.6.1

The values of $\Omega_l = \xi_1 - \xi_2 T$ in Eq. (4.6.1) for the liquid mixing enthalpy parameter Ω_l in cal/mole and other parameters in the liquid phase.

System	Ω_l (cal/mole)	ΔS_m (eu)	T_m (C)
AlP	2800-4.80 T	15.0	2530
AlAs	600-12 T	15.6	1770
AlSb	1230-10 T	14.74	1065
GaP	28000-4.8 T	16.8	1465
GaAs	5160-0.16 T	16.64	1238
GaSb	4700-6 T	15.8	710
InP	4500-4 T	14.0	1070
InAs	3860-10 T	14.52	942
InSb	3400-12 T	14.32	525
Al-Ga	104		
Al-In	1060		
Ga-In	1060		
P-As	1500		
As-Sb	750		

After Stringfellow (1974)

Table 4.6.2

Mixing energy parameters Ω in Kcal/mole. For models in which the mixing enthalpy ΔH is a function of both x and T , Ω is defined as $\Delta H = x(1-x)\Omega$ for a random alloy ($T \rightarrow \infty$) at $x=0.5$.

system	DLP ^a	FM ^b	IMP ^c	MZ ^d	CS ^e	GQCA ^f	WF ^g	Exp		
(Ga,Al)P			0.00		-0.05	0.01				
(Ga,Al)As	0.02	0.03	0.00	0.02	-0.07	0.03	0.30	0.0 ^h		
(Ga,Al)Sb	0.02	0.03	0.02	0.02	-0.15	0.12		0.0 ^h		
(Ga,In)P	3.63	2.94	3.00	4.56	2.54	2.85	3.07	3.25 ⁱ	3.40 ^k	3.50 ^h
(Ga,In)As	2.81	2.42	2.36	2.49	1.60	2.19	2.35	1.65 ^j	2.00 ^j	3.00 ^h
(Ga,In)Sb	1.85	1.83	1.77	2.53	0.81	1.72		1.48 ^j	1.90 ^h	
(In,Al)P			2.77		2.55	2.66				
(In,Al)As	2.81	2.37	2.32	3.60	2.17	2.26			2.50 ^h	
(In,Al)Sb	1.46	1.45	1.49	2.06	1.36	1.46			0.60 ^h	
(Cd,Zn)Te	1.97	1.63	1.73	2.12	1.24	1.59	2.29	1.34 ^m		
(Hg,Cd)Te			0.00		-0.07	0.03	0.38	0.72 ⁿ	1.40 ^m	
(Hg,Zn)Te	1.81	1.48	1.56	1.91	1.50	1.55	1.88	3.00 ^m		
Al(P,As)			0.65		0.76	0.67				
Ga(P,As)	0.98	0.66	0.70	1.15	0.94	0.74	0.91	0.40 ^h	1.00 ^l	1.26 ^p
In(P,As)	0.58	0.52	0.52	0.72	0.57	0.52		0.40 ^h		
Al(As,Sb)			3.38		4.09	3.45				
Ga(As,Sb)	3.35	2.76	2.81	4.38	3.67	2.92		4.00 ^k	4.27 ^p	4.50 ^h
In(As,Sb)	2.29	2.17	2.23	2.89	2.52	2.20		2.25 ^h	2.90 ^j	
Al(P,Sb)			6.99		8.32	6.94				
Ga(P,Sb)			6.36		8.66	6.57				
In(P,Sb)			5.08		5.76	4.97				
Zn(S,Se)			0.98		0.90	0.95				
Zn(Se,Te)	3.11	2.12	2.23	2.91	2.26	2.16		1.55		
Zn(S,Te)			6.45		6.20	6.12				

a: Stringfellow (1974), Ω_{DLP} in Eq.(4.6.5).

b: Fedders and Muller (1984), Ω_{FM} in Eq.(4.6.7).

c: Chen and Sher (1985), Ω_{IMP} in Eq.(4.6.13).

d: Martins and Zunger (1984), VFF model.

e: Chen and Sher (1985), VFF plus BOM perturbation.

f: Sher et al. (1987), the 16-bond model in Sec. 4.9.

g: Wei et al. (1990), LDA calculation described in Sec. 4.11.

h: Panish and Ilgemes (1972).

i: Foster (1972).

j: Antypas (1970b).

k: Foster and Woods (1971).

l: Antypas (1970a).

m: Laugies (1973).

n: Su et al. (1985), value taken at 1000K.

p: Stringfellow (1974, 1982).

4.6.1

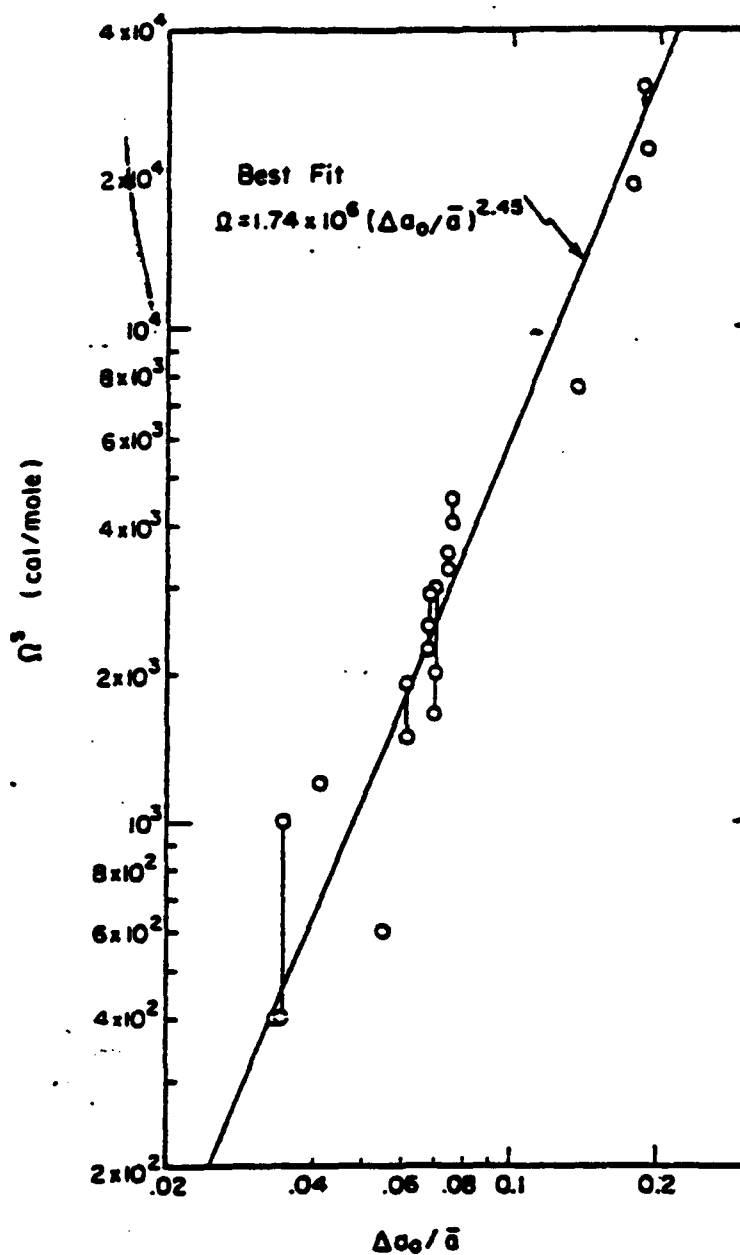


Figure 4.6.1 Log-log plot of experimental values of Ω versus percentage lattice constant difference. The solid line represents the best straight-line fit. After Stringfellow (1974).

Stringfellow (1974). The best straight line fit gives a slope of 2.45, yielding the following fitted function:

$$\Omega_{\text{fit}} = 1.174 \times 10^6 (\Delta a/a)^{2.45} , \quad (4.6.3)$$

where $a = (a_{\text{AC}} + a_{\text{BC}})/2$. This result prompted Stringfellow to correlate with Phillips and Van Vechtan's (1970) dielectric model, which relates the alloy formation energy to the energy-gap parameter E_h [see Eq. (2.6.4)], which in turn is roughly proportional to $a^{-2.5}$. By simply assuming that the formation energy has the form

$$E = -\kappa/a^{2.5} \quad (4.6.4)$$

for every semiconductor, then the mixing enthalpy is given by

$$\begin{aligned} \Delta H &= E(\text{alloy}) - xE(\text{BC}) - (1-x) E(\text{AC}) \\ &= -\kappa [a_{\text{alloy}}^{-2.5} - x a_{\text{BC}}^{-2.5} - (1-x) a_{\text{AC}}^{-2.5}] \\ &= 4.375 \kappa x (1-x) (\Delta a)^2 / a^{4.5} = x(1-x) \Omega_{\text{DLP}} \end{aligned} \quad (4.6.5)$$

In Eq. (4.6.5) the alloy lattice constant a_{alloy} is assumed to be the concentration weighted average (Vegard's Law), and only terms to second order in Δa in the Taylor expansion have been kept. If κ is treated as an adjustable parameter, the best fit to the III-V alloy data yields $\kappa = 1.15$ for Ω in kcal/mol and Δa and a in Å. This model gives very good results, as indicated in Table 4.6.2.

Arguing from a different point of view Fedders and Muller (1984) suggested that the Δa^2 dependence in Eq. (4.6.5) might originate from strain energy. If the alloying atoms A and B are held rigidly on their fcc sublattice sites, and if the alloy lattice constant is taken to be the concentration weighted average (Vegard's law), then the strain energy in the alloy can be shown to be $\Delta E = x(1-x)\Omega_{\text{FM}}$ with a mixing energy parameter given by

$$\Omega_{\text{FM}} = \frac{9}{2} B V (\Delta a/a)^2 , \quad (4.6.6)$$

where B is the average bulk modulus and V the average molar volume. Because B in semiconductors varies as a^{-n} with n ranging from 5 to 9 (see

Chapter 3), Ω_{FM} is seen to behave like $(\Delta a)^2/a^{n-1}$. Thus the power law dependences on Δa and a in Ω_{DLP} and Ω_{FM} are similar. However, Ω_{FM} is about four times larger than the experimental values. In fact a scaled down strain contribution,

$$\Omega_{scale} = 0.226 \Omega_{FM} \quad (4.6.7)$$

yields results comparable to those of Ω_{DLP} . The reduction of Ω_{FM} is mainly due to bond length relaxation neglected in the above derivation.

If strain energy is the main contribution to the mixing enthalpy, then the valence force field model (VFF) given in Eq. (3.3.1) should provide a means to formulate a systematic calculation. One can replace an A atom in the AC crystal by an impurity B atom and calculate the substitution energy $\Delta E(B \text{ in } AC)$ and vice versa for $\Delta E(A \text{ in } BC)$, then the mixing enthalpy parameter is given by (Chen and Sher, 1985)

$$\Omega_{imp} = [\Delta E(A \text{ in } BC) + \Delta E(B \text{ in } AC)] / 2 . \quad (4.6.8)$$

To calculate these energies precisely in VFF, one has to allow the atoms in the first several shells about the impurity to relax. The results of such calculation (Martins and Zunger, 1984; Chen and Sher, 1985) do give good results for mixing energy and bond lengths between the impurity and the first shell host atoms. However, it was found (Chen and Sher, 1985) that by neglecting the bond angle terms in the VFF ($\beta = 0$) and at the same time truncating the range of the atomic relaxation at the second shell, a simple spring model similar to that shown in Figure 3.7.1 yields nearly the same results as the extended VFF for both the impurity bond lengths and alloy mixing energy. It is instructive to examine this simple spring model closely. Let the sixteen bonds before the substitution be the host BC bonds with VFF force constant α and bond length d . When the central B atom is removed and an A atom is substituted in its place, the bond length d_1 of the first shell and d_2 of the second shell are different from d . The strain energy in VFF with $\beta = 0$, and assuming the second shell atoms are held firmly in place, is then given by

$$\Delta E(A \text{ in } BC) = 4 \left[\frac{3}{2} \alpha_1 (d_1 - d)^2 \right] + 12 \left[\frac{3}{2} \alpha (d_2 - d)^2 \right] \quad (4.6.9)$$

where the subscript I denotes the impurity bond in its own crystal (i.e., the AC bond in the pure AC crystal). Let $\delta = (d_1 - d)/d$ and $\delta_0 = (d_I - d)/d$, then in the relaxed configuration $(d_2 - d)/d = -\delta/3$ and Eq. (4.6.9) becomes

$$\Delta E(\text{A in BC}) = 4 \left[\frac{3}{2} \alpha_I (\delta - \delta_0)^2 + \frac{1}{2} \alpha \delta^2 \right] d^2 \quad (4.6.10)$$

Minimizing ΔE with respect to δ yields the following value of d at the equilibrium position

$$\delta = \delta_0 / \left(1 + \frac{1}{3} \frac{\alpha}{\alpha_I} \right). \quad (4.6.11)$$

If both α and α_I are nearly equal then

$$\delta \approx \frac{3}{4} \delta_0 \quad (4.6.12)$$

and $\Delta E(\text{A in BC}) = 3\alpha\delta_0^2 d^2 / 2$. Similar results can be obtained for $\Delta E(\text{B in an AC host})$. Equation (4.6.9) then yields the following estimate for the mixing energy parameter from this impurity model,

$$\Omega_{\text{imp}} = 3\alpha\delta_0^2 d^2 / 2 \quad (4.6.13)$$

Note that this model, Ω_{imp} is precisely one quarter of Ω_{FM} . Thus this simple relaxed lattice model not only explains why the strain energy is reduced by 1/4 but also gives a very good prediction of the value $\gamma = 0.75$ for the lattice relaxation [compare Eq. (4.6.12) with the definition of γ in Eq. (1.3.1) and Table 1.3.1] found in pseudobinary alloys from the EXAFS experiments. While this short range strain model with β set to zero produces accurate results, keep in mind that it is not the way nature actually behaves. The approximation works only because effects from long range relaxation nearly cancel the lowest order modifications caused by the β terms. If one asks questions that involve shears, or for more detailed structural information, e.g., the second neighbor distances, then effects due to β and long range relaxations must be included.

4.7 Generalized Quasi-Chemical Theory (GQCA)

There are at least two reasons to extend the theory beyond the first approximation. First, the effects of the neglected statistical correlations should be improved. Second, and perhaps more importantly, the microscopic energies that govern the statistics may not be described by

simple pair energies. In this section, QCA is generalized to deal with clusters of any size.

Let each cluster contain n atoms on the sublattice occupied by the A and B atoms and let $J+1$ be the number of different configurations in which the A and B atoms can be arranged on a cluster with a distinct energy ϵ_j , for $j = 0, 1, \dots, J$. For example, in the case of a tetrahedral cluster in which ϵ_j differs only when the number of B atoms on the cluster changes, then the index j will equal the number of B atoms on the cluster, and therefore takes on five values $j = 0, 1, 2, 3$, and 4 . However, for a square with two A atoms and two B atoms on the corners the energy for two atoms on the same side may be different from that when they are on a diagonal. So J may be larger than n . Obviously, the degeneracies of arrangements with the same ϵ_j tend to decrease if $J > n$ since the total number of configurations is fixed, i.e., for a 4 site case, it is 2^4 . Let M be the total number of clusters in the alloy, M_j be the number of the j -type clusters so that $M = \sum M_j$, and $x_j = M_j/M$ the fraction of the j -type clusters. Let us further define $n_j^A(n_j^B)$ to be the number of A (B) atoms in a j -type cluster, and the notation n_j will stand for n_j^B . Since the concentration x is fixed, the set of fractions x_j are constrained to satisfy

$$nx = \sum_{j=0}^J n_j^B x_j . \quad (4.7.1)$$

The mixing energy for a specified set $\{M_j\}$ is then given by

$$\Delta E = \sum_{j=0}^J M_j \epsilon_j - M[(1-x) \epsilon_A^0(n) + x \epsilon_B^0(n)] , \quad (4.7.2)$$

where $\epsilon_A^0(n)$ and $\epsilon_B^0(n)$ are the energies of an n -atom cluster in the pure AC and pure BC crystals respectively, as indicated by the superscript 0. We note that the cluster energy $\epsilon_0 = \epsilon_A(n)$, which is the energy in the alloy for a cluster containing all A atoms, in general differs from $\epsilon_A^0(n)$, because of the influence of strain and different chemical bonding. Let $\epsilon_A = \epsilon_0 - \epsilon_A^0(n)$ and similarly for $\epsilon_B = \epsilon_J - \epsilon_B^0(n)$. Then Eq. (4.7.2) can be written as

$$\Delta E = M[(1-x) \epsilon_A + x \epsilon_B] + M \sum_j x_j \Delta_j , \quad (4.7.3)$$

where the reduced excess energies Δ_j are defined as

$$\Delta_j = \epsilon_j - \frac{n - n_j}{n} \epsilon_0 - \frac{n_j}{n} \epsilon_J , \quad (4.7.4)$$

and use has been made of Eq. (4.7.1) Note that by definition Δ_0 and Δ_J are zero.

All the temperature dependence arising from the statistical mechanics is, as we shall see, contained in the x_j factors. Therefore, the first term of ΔE in Eq. (4.7.3) does not influence the temperature dependence arising from averaged statistical quantities for a given x ; nearly all the temperature dependence is in the small second term. The mixing entropy can be calculated from $\ln \Phi$ with Φ given by [an extension of Eq. (4.2.9)]

$$\Phi = \frac{N!}{N_A! N_B!} \frac{M!}{\prod_j M_j!} \prod_j (x_j^0)^{M_j} , \quad (4.7.5)$$

where x_j^0 is the *a priori* probability, i.e., the random-alloy value

$$x_j^0 = g_j (1-x)^{(n-n_j)} x^{n_j} , \quad (4.7.6)$$

and g_j is the degeneracy of clusters with energy ϵ_j . The resulting mixing entropy takes a simple form

$$\Delta S = -Nk (x \ln x + y \ln y) - Mk \sum_j x_j \ln (x_j / x_j^0) . \quad (4.7.7)$$

Eqs. (4.7.2) and (4.7.7) thus combine to give an explicit expression for the mixing free energy $\Delta F = \Delta E - T\Delta S$ in terms of x_j .

To find the equilibrium cluster probability distribution $\{\bar{x}_j\}$, one takes the partial derivatives $\partial \Delta F / \partial x_j$ and sets them to zero. Although there are $J + 1$ unknown x_j 's, there are two independent constraints, one given by Eq. (4.7.1) and the other is that the total probability is unity,

$$\sum_j x_j = 1 . \quad (4.7.8)$$

This leaves us with $J - 1$ nonlinear coupled equations to be solved for the thermally averaged $\{\bar{x}_j\}$. While the set $\{\bar{x}_j\}$ are the equilibrium values of

the more general set $\{x_j\}$, we shall leave the bar off since it causes no confusion. A simple approach to the numerical problem is to use the Lagrange multiplier formalism in the constrained variational calculation for the $\{x_j\}$ set, i.e.,

$$\frac{\partial}{\partial x_j} \left\{ \Delta F / M - \gamma \left[\left(\sum_j x_j \right) - 1 \right] - \alpha \left[\left(\sum_j n_j x_j \right) - nx \right] \right\} = 0 . \quad (4.7.9)$$

This relates x_j to the Lagrange multipliers γ and α

$$x_j = x_j^0 \exp [(\gamma + \alpha n_j - \Delta_j) / kT] . \quad (4.7.10)$$

The normalization condition (4.7.8) is then used to eliminate γ and obtain the following result

$$x_j = x_j^0 \exp [(\alpha n_j - \Delta_j) / kT] / \sum_j x_j^0 \exp [(\alpha n_j - \Delta_j) / kT] . \quad (4.7.11)$$

The algebra is simplified by defining a new variable $\eta = x e^{\alpha \beta} / y$. Then

$$x_j = g_j \eta^{n_j} e^{-\beta \Delta_j} / Z \quad (4.7.12)$$

where $\beta = 1/(kT)$, and Z is a cluster grand partition function

$$Z = \sum_j g_j \eta^{n_j} e^{-\beta \Delta_j} \quad (4.7.13)$$

In general the thermal average of any dynamic variable D is

$$\bar{D} = \sum_j D_j x_j \quad (4.7.14)$$

Then the expression for x_j in Eq. (4.7.12) is used in the other constraint equation (4.7.1) to arrive at the following n^{th} order polynomial equation for the unknown parameter η

$$\bar{n}_j = \sum_j n_j x_j = \sum_j n_j g_j \eta^{n_j} e^{-\beta \Delta_j} / Z = nx \quad (4.7.15)$$

We shall refer to the above argument as the generalized QCA (GQCA), while QCA is reserved for the pair interaction case.

To illustrate the usefulness of the equations developed in this section, let us consider the case in which the cluster energies ϵ_j depend only on the number of A and B atoms in the cluster. For a cluster containing n atoms there are a total of $n + 1$ distinct energies of the cluster type $A_{(n-j)} B_{(j)}$ with j ranging from 0 to n and with $n_j = j$. Then the cluster degeneracy is simply

$$g_j = \frac{n!}{j! (n-j)!} \quad (4.7.16)$$

In terms of η and Z the mixing free energy is given by

$$\Delta F = N\ell (y\epsilon_A + x\epsilon_B) + NkT \left[(1 - n\ell) (x \ln x + y \ln y) - \ell \ln Z + n\ell x \ln \eta \right], \quad (4.7.17)$$

where we have defined $\ell = M/N$. Notice that a term from ΔS has exactly cancelled the temperature dependent term $\sum_j x_j \Delta_j$ in the enthalpy. Thus it is impossible to determine the temperature variation of ΔE from a measurement of ΔF .

To determine the miscibility gap, the spinodal curve, and critical temperature we need to examine the derivatives of ΔF with respect to x . They are simply given by

$$\begin{aligned} \frac{\partial \Delta F}{\partial x} = N\ell \left(\epsilon_B - \epsilon_A + y \frac{\partial \epsilon_A}{\partial x} + x \frac{\partial \epsilon_B}{\partial x} \right) \\ + NkT \left[(1 - n\ell) \ln (x/y) + \frac{\ell}{kT} \left(\frac{\partial \Delta_j}{\partial x} \right) + n\ell \ln \eta \right] \end{aligned} \quad (4.7.18)$$

and

$$\begin{aligned} \frac{\partial^2 \Delta F}{\partial x^2} = N\ell \left[2 \left(\frac{\partial \epsilon_B}{\partial x} - \frac{\partial \epsilon_A}{\partial x} \right) + y \frac{\partial^2 \epsilon_A}{\partial x^2} + x \frac{\partial^2 \epsilon_B}{\partial x^2} \right] \\ + NkT \left[(1 - n\ell) / (xy) + \left(\frac{\ell}{kT} \right) \frac{\partial}{\partial x} \left(\frac{\partial \Delta_j}{\partial x} \right) + n\ell \frac{\partial \ln \eta}{\partial x} \right] \end{aligned} \quad (4.7.19)$$

Equation (4.7.17) is obtained from Eq. (4.7.16) by taking the partial of Z given by Eq. (4.7.13) with respect to x , and combining the result with Eqs. (4.7.12) and (4.7.14) to yield the relation

$$nx \frac{\partial \ln \eta}{\partial x} = \frac{\partial \ln Z}{\partial x} + \frac{1}{kT} \left(\frac{\partial \Delta_j}{\partial x} \right). \quad (4.7.20)$$

If the $\partial/\partial x$ is also taken of Eq. (4.7.15), that produces a second equation in $\partial \ln \eta / \partial x$ and $\partial \ln Z / \partial x$ which can be solved along with Eq. (4.7.20)

$$\frac{\partial \ln \eta}{\partial x} = \frac{n + \frac{1}{kT} \left[\left(n_j \frac{\partial \Delta_j}{\partial x} \right) - nx \left(\frac{\partial \Delta_j}{\partial x} \right) \right]}{\bar{n}_j^2 - n^2 x^2} - \beta \left(\frac{\partial \Delta_j}{\partial x} \right), \quad (4.7.21)$$

and

$$\frac{\partial \ln Z}{\partial x} = nx \left[\frac{n + \frac{1}{kT} \left(n_j \frac{\partial \Delta_j}{\partial x} \right) - \frac{nx}{kT} \left(\frac{\partial \Delta_j}{\partial x} \right)}{\bar{n}_j^2 - n^2 x^2} \right] - \beta \left(\frac{\partial \Delta_j}{\partial x} \right). \quad (4.7.22)$$

Equations (4.7.18 - 4.7.22) recognize that the excess energies ϵ_A , ϵ_B and the set $\{\Delta_j\}$ have an explicit x dependence. This can be seen in Fig. (4.9.2) where this x dependence for $G_{A1-x} \ln_x A_s$ is plotted. In this case the variation of the set $\{\Delta_j\}$ is quite small. However, as we shall demonstrate later for lattice mismatched alloys $\epsilon_A \equiv \epsilon_{A0} x^2$ and $\epsilon_B \equiv \epsilon_{B0} y^2$ are good approximations where $\epsilon_{A0} \equiv \epsilon_{B0} \gg \Delta_j$ for all j . This dependence arises because a pure A or pure B cluster has large strain energy in the alloy whose average lattice constant follows Vegard's law (see Fig. 1.5). Given these x dependences of ϵ_A and ϵ_B the leading mixing enthalpy term in Eq. (4.7.7) becomes

$$N \Omega (y \epsilon_A + x \epsilon_B) = N \Omega' x(1-x) + N \Omega (\epsilon_{B0} - \epsilon_{A0}) x^2(1-x)$$

where

$$\Omega' \equiv \Omega \epsilon_{B0}.$$

Most often one finds $\epsilon_{B0} - \epsilon_{A0} \ll \epsilon_{B0}$ so the mixing enthalpy has the usual form $N\Omega' x(1-x)$ that was assumed in Section 4.2, Eq. 4.2.3.

Equations (4.7.18 - 4.7.22) simplify greatly if the $\frac{1}{kT} \left(\frac{\partial \epsilon_j}{\partial x} \right)$ terms can be neglected. There are some exceptional cases where g_j may also have a slow x dependence but they are so rare that we have not included this possibility in Eqs. (4.7.17 - 4.7.22).

Note that once η is deduced by solving Eq. (4.7.15), $\partial\eta/\partial x$ can be found from Eq. (4.7.21). Then the miscibility gap, the spinodal curve, and the critical temperature can be evaluated.

If Δ_j contains only pair potential combinations, then ΔF is a symmetric function of x about $x = 1/2$ for a given T . The miscibility critical points x_1 and x_2 , if they exist at a given T , are determined by the condition $\partial\Delta F/\partial x = 0$ and are related by $x_1 = (1-x_2)$. Similarly, the values of x'_1 and x'_2 of the spinodal are determined by the condition $\partial^2\Delta F/\partial x^2 = 0$ and are also related by $x'_1 = (1-x'_2)$. The critical temperature T_c arises from the special case where $x_1 = x_2 = x'_1 = x'_2 = 1/2$ and $\partial^2\Delta F/\partial x^2 = 0$. At $x = 1/2$, the solution for η is $\eta=1$. This result can be obtained by inspecting Eq. (4.7.15) and knowing that interchanging the A and B atoms does not change the number of AB pairs in a cluster so $\Delta_j = \Delta_{n-j}$, and also $g_j = g_{n-j}$. Using this η value, one can express $\partial \ln \eta / \partial x$ at $x = 1/2$ explicitly in terms of T . Eqs. (4.7.19) and (4.7.21) then lead to the following explicit equation for the critical temperature T_c .

$$-2 \epsilon_B^0 / kT + n^2 \left(\frac{1 - n\ell}{\ell} \right) + \frac{4}{n_j^2 - \frac{n^2}{4}} = 0 \quad (4.7.23)$$

where for simplicity $(\epsilon_{B0} - \epsilon_{A0})/kT$ and $\frac{1}{kT} \frac{\partial \epsilon_j}{\partial x}$ have to be set to zero.

For the case in which the clusters are pairs, i.e., QCA, we have $n = 2$, and $q = z/2$, where z is the coordination number. Then Eq. (4.7.23) reduces to the following simple equation

$$\Delta_1 / kT_c = - \ln \left(\frac{\epsilon_{B0}}{2kT} + \frac{z-2}{z} \right) \quad (4.7.24)$$

Comparing Eqs. (4.7.23) and Eq. (4.2.16) we see they are superficially similar, but in fact are quite different. The reason for the difference is that the x dependence of the $\{\Delta_j\}$ set is not properly treated in the derivation of Eq. (4.2.16). The ratio of the critical temperature to Δ_1 for different ϵ_{B0}/Δ_1 ratios are in Table 4.7.1.

Table 4.7.1

Critical temperature for different coordination numbers z , and ϵ_{B0}/Δ_1 ratios.

z	4	12
ϵ_{B0}/Δ_1	kT_c/Δ_1	kT_c/Δ_1
0	1.44	5.49
5	6.90	20.8
10	11.9	36.0
15	16.9	51.3

As expected, as the ratio ϵ_{B0}/Δ_1 increases T_c increases, because ϵ_{B0} becomes the proper reference energy.

It is common to find ratios $\epsilon_{B0} / \Delta_1 \approx 10$ so theories that do not account properly for the x dependence of the energies are inaccurate.

The chemical potential μ_B for B atoms in the $A_{1-x}B_x$ or $A_{1-x}B_xC$ alloy systems is defined as

$$\mu_B \equiv \frac{\partial \Delta F}{\partial N_B} = \frac{\partial \Delta F}{N \partial x} \quad (4.7.25)$$

Then from Eq. (4.7.18) and using the approximation $\frac{\epsilon_{A0} - \epsilon_{B0}}{kT}, \frac{1}{kT} \frac{\partial \Delta_j}{\partial x} \ll 1$, the leading terms are

$$\mu_B = (1-2x) \epsilon_{B0} + n\ell \mu_{BR} + kT (1-n\ell) \ln(x/y) \quad (4.7.26)$$

where the reduced chemical potential μ_{BR} is related to the reduced activity coefficient η by

$$\mu_{BR} \equiv kT \ln \eta. \quad (4.7.27)$$

The quantity μ_{BR} is the contribution to the chemical potential stemming from the "state of order" of the alloy.

To gain some insight into the behavior of μ_{BR} examine Eq. (4.7.15) in the case where $J = n$ and $n_j = j$. Expand Δ_j in a power series in j

$$\Delta_j = \sum_i \delta_i j^i \quad (4.7.28)$$

As we will see in Sec. 4.9 it is often sufficient to retain terms in this expansion to j^2 . To allow us to obtain a simple analytical form that displays the major physical trends of μ_{BR} we will also replace $j^2 \approx \bar{j}j$. Then Eqs. (4.7.28) and (4.7.15) become

$$\begin{aligned} nx &= e^{-\delta_0/kT} \sum_{j=0}^n j g_1 [\eta e^{-(\delta_1 + \delta_2 j)/kT}]^j / Z \\ &= \frac{n \eta e^{-(\delta_1 + nx \delta_2)/kT}}{1 + \eta e^{-(\delta_1 + nx \delta_2)/kT}} \end{aligned} \quad (4.7.29)$$

or

$$\mu_{BR} \equiv (\delta_1 + nx \delta_2) + kT \ln x/y. \quad (4.7.30)$$

Therefore, μ_{BR} is of order $\bar{\Delta}_j$ and has an entropy contribution that dominates at high temperatures. This entropy contribution exactly cancels the $kT n \ell \ln(x/y)$ term in Eq. (4.7.26) so it simplifies to

$$\mu_B \equiv \ell (1-2x) \epsilon_{B0} = n\ell (\delta_1 + nx \delta_2) + kT \ln (x/y). \quad (4.7.31)$$

The entropy contribution to the chemical potential is only that for a random alloy. All reference to modifications caused by GQCA have cancelled out. This result is nearly exact in the temperature limit high

enough so $\{\Delta_j\} \ll kT \ln(x/y)$, because in this limit none of the approximations used to obtain Eq. (4.7.29) are needed.

Note that the cluster distribution function in Eq. (4.7.12) could have been written directly had we used a simple grand canonical ensemble for the clusters from the outset. Then η is the absolute activity for the B atoms in this ensemble. The whole derivation can also be done using a steepest descents argument for the partition function (Sher et al., 1987). While these methods are more elegant, the essential physics is the same as the GQCA derived in the straightforward formalism above. These calculations can be summarized in four steps: (1) calculate the cluster energies $\epsilon_j(\eta)$ and from them the reduced excess energies Δ_j defined in Eq. (4.7.4); (2) solve the polynomial equation (4.7.14) to obtain η ; (3) use η and the energies to obtain the cluster distribution from Eq. (4.7.12) and use this distribution to calculate all the statistical averaged properties for the alloy; and (4) in particular, use the set $\{x_j\}$ in Eq. (4.7.3) to obtain the mixing enthalpy ΔE , and in Eq. (4.7.7) to obtain ΔS , which then enables phase-diagram calculations and thermodynamic studies.

4.8 Internal Strain and Cluster Energies

Strictly speaking, QCA and GQCA imply short-ranged interaction energies. However, the energy models in Sec. 4.6, which are consistent with the experiments, indicate that strain is the most important contribution to mixing energy in most pseudo-binary alloys. The exceptions are the near lattice constant matched alloys, e.g., GaAlAs, and HgCdTe. Since the strain energy is shared among bonds throughout the crystal and is mutually interactive in nature, it can never be separated into isolated contributions. To write the mixing enthalpy as a sum of cluster energies, these cluster energies must contain contributions extending beyond the cluster size. Such an effect has already been included in the energy models used in Section 4.6. This section is designed to provide a basis for treating the internal strain energy in alloys (Chen et al., 1987) and obtaining its contribution to cluster energies for insertion into the GQCA calculation.

Consider an alloy divided into M nonoverlapping clusters. In general, alloy strain energy can be written as the sum of the strain energies of these clusters, and the interaction among the clusters. Focus on a particular cluster denoted n , detach this cluster from the alloy, and allow both the cluster and the remaining medium to relax. When the cluster is reconnected to the medium, there will be a change in the energy. This energy change denoted h_{nm} is defined as the interaction energy between the cluster and its surrounding medium labeled by m . This the total strain energy of the alloy can be written as

$$E = \sum_n \left(\epsilon_n + \frac{1}{2} h_{nm} \right). \quad (4.8.1)$$

Here ϵ_n is a residual strain energy for the n^{th} cluster when it is detached from the alloy and is in its equilibrium configuration. The sum in Eq. (4.8.1) is over all clusters in the alloy. The factor $1/2$ is inserted to eliminate double counting. In an alloy, Eq. (4.8.1) can be replaced by an ensemble average.

$$E = M \sum_n \sum_m \left(\epsilon_n + \frac{1}{2} h_{nm} \right) p_n P_{nm}, \quad (4.8.2)$$

where p_n is the probability that a cluster will be of the n type (e.g., specified by the number of A and B atoms and their arrangements) and P_{nm} is a conditional probability that the surrounding environment is in state m when the cluster is in state n . Note that in Eq. (4.8.1) n denotes the location of the cluster, while in Eq. (4.8.2) n designates the state of a cluster. From Eq. (4.8.2), an effective cluster energy $\epsilon(n)$ can be defined as

$$\epsilon(n) = \sum_m \left(\epsilon_n + \frac{1}{2} h_{nm} \right) P_{nm}. \quad (4.8.3)$$

These energies can be expressed explicitly within the following elastic continuum model.

Consider a spherical cluster embedded in an elastic medium. The interaction energy can be written as

$$h_{nm} = \frac{1}{2} \mu_n (a_n - r)^2 + \frac{1}{2} v_m (b_m - r)^2, \quad (4.8.4)$$

where the equilibrium radius r is to be determined by minimizing h_{nm} . The lengths a_n and b_m are the "natural" radii and μ_n and v_m are the force

constants of the cluster and the medium respectively. For a macroscopic semiconductor cluster of radius R , $\mu = 12\pi BR$ and a good approximation for v is $v = 3.2\pi(C_{11} - C_{12} + 3C_{44})$ (Chen and Sher, 1985). The value of r corresponding to the minimum value of h_{nm} is given by

$$r_{nm} = (\mu_n a_n + v_m b_m) / (\mu_n + v_m) , \quad (4.8.5)$$

and the actual interaction energy becomes

$$h_{nm} = \frac{1}{2} \mu_n v_m (a_n - b_m)^2 / (\mu_n + v_m) . \quad (4.8.6)$$

Next cast the average energy into a form in which a_n and b_m are decoupled. To do this, an effective radius r_e is introduced such that

$$\begin{aligned} & \sum_n \sum_m p_n p_m \mu_n v_m (a_n - b_m)^2 / (\mu_n + v_m) \\ &= \sum_n \sum_m p_n p_m \mu_n v_m [(a_n - r_e)^2 + (b_m - r_e)^2] / (\mu_n + v_m), \end{aligned} \quad (4.8.7)$$

which imposes the following condition on r_e :

$$\sum_n \sum_m p_n p_m \mu_n v_m (a_n - r_e)^2 (b_m - r_e)^2 / (\mu_n + v_m) = 0. \quad (4.8.8)$$

A very quick and important result for the internal strain energy can be obtained from these two equations by neglecting the correlation between p_n and P_{nm} and approximating all the force constants by the concentration weighted average values μ and v . This result is: $r_e = \langle a_n \rangle = \langle b_m \rangle$, $\langle (a_n - r_e)^2 \rangle = \langle (b_m - r_e)^2 \rangle = \langle \Delta r^2 \rangle$ and

$$E = M \left[\epsilon + \frac{1}{2} \mu v \langle \Delta r^2 \rangle / (\mu + v) \right] , \quad (4.8.9)$$

where ϵ is the average value $\langle \epsilon_n \rangle$.

Eq. (4.8.9) shows that there are three essential features contributing to the internal strain energy in this cluster theory: the mean square fluctuation of the cluster radii $\langle \Delta r^2 \rangle$, the effective force constants which are $\mu v / (\mu + v)$, and the mean strain energy ϵ of isolated clusters.

This result further suggests a way to use the effective medium to obtain cluster energies in an alloy. For example, if one assumes that the effective medium has an effective elastic force constant v and radius r_e , then replacing the medium by a cluster n introduces a strain energy having the same expression as Eq. (4.8.6) with v_m replaced by v and b_m by r_e . If all the strain energy is assigned to this cluster, we obtain an effective cluster energy,

$$\epsilon_c(n) = \epsilon_n + \frac{1}{2} \mu_n v (a_n - r_e)^2 / (\mu_n + v) . \quad (4.8.10)$$

We can see that the alloy mean energy $E = M \sum p_n \epsilon_c(n)$ reduces to the result in Eq. (4.8.9) when the μ_n is taken to be the mean value μ . The effective medium nature enters if we take r_e to be determined by the condition that the average energy E be a minimum, i.e.,

$$\sum_n p_n \mu_n v (a_n - r_e) / (\mu_n + v) = 0 . \quad (4.8.11)$$

Application of this self-consistent approach to the elastic constants has already been considered in Section 3.7.

Let us now apply the general ideas behind Eqs. (4.8.10) and (4.8.11) for the cluster energies, and GQCA to a systematic discussion of the statistical properties of pseudobinary alloys $A_{1-x}B_xC$. Because of the special atomic arrangements of these alloys, the bonds are the natural entities to use for the energy counting while it is more convenient to count the atoms for the statistical arrangements. All the bonds in these alloys can be clearly separated into AC and BC bonds. A convenient way to generate the cluster is to assign the four AC bonds surrounding an A atom to that atom, and similarly assign the four BC bonds to the central B atom. In this way the smallest cluster is a one-atom four-bond cluster. Following the ideas behind Eq. (4.8.10) we start with an effective medium and replace the

center atom by an A atom, the four surrounding bonds by AC bonds, and calculate the interaction energy to obtain the one-atom cluster energy ϵ_A . In a similar way we can obtain ϵ_B by replacing the cluster with a B atom and its four BC bonds. Note that the mixing energy ΔE in Eq. (4.7.3) is not zero, because ϵ_A and ϵ_B in an effective medium are different from their respective values ϵ_A^0 and ϵ_B^0 in their respective pure constituent crystals. If the effective spring model of Figure 3.7.3 is used, then $\mu = 6\alpha_A$, $\nu = 2\alpha$, and we have [following Eqs. (4.8.10) and (4.7.4)]

$$\Delta_A = \frac{3\alpha_A\alpha}{3\alpha_A + \alpha} (d_{AC}^0 - d)^2 = \frac{3\alpha_A\alpha}{3\alpha_A + \alpha} (1-x)^2 \delta_0^2 d^2, \quad (4.8.12)$$

where δ_0 is $(d_{AC}^0 - d_{BC}^0) / d$ with d being the VCA bond length. A similar expression can be obtained for Δ_B . When d and all the α 's in Δ_A and Δ_B are replaced by their mean values, the resulting ΔE is only modified by a few percent from those based on the actual expressions in Eq. (4.8.12). The resulting mixing enthalpy is given by a simple expression $\Delta E = x(1-x) 3\delta_0^2 d^2 / 2$. This is the same result as that given in Eq. (4.6.13) for the mixing enthalpy parameter. Since each cluster contains only one alloying atom and none of the clusters share the same alloying atom, the total possible ways to arrange the clusters is exactly Φ_0 in Eq. (4.2.1), so that the total mixing free energy in this model is precisely Stringfellow's preferred form given in Eq. (4.6.2).

The next simplest cluster set for the pseudobinary alloy are clusters that contain four alloy atoms and sixteen bonds as depicted in Figure 4.8.1. This is the smallest cluster size that can provide a meaningful

fig 4.8.

A_2B_2 16-Bond Cluster

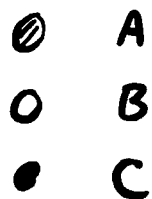
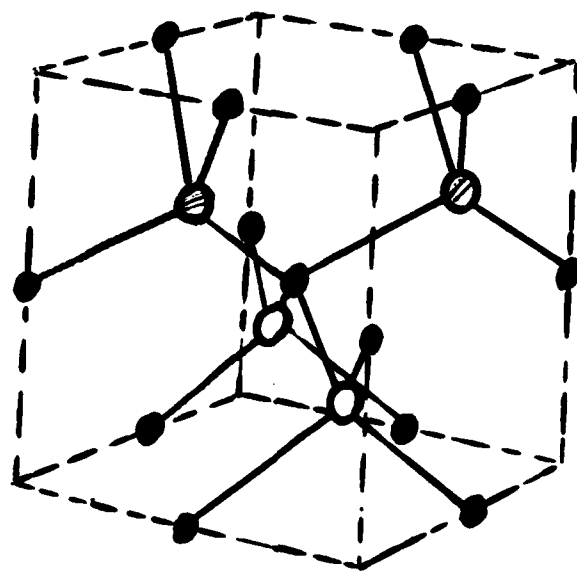


Figure 4.8.1 Schematic picture of a 4-atom, 16-bond cluster in a pseudobinary alloy. This figure shows an A_2B_2 cluster in a cubic cage.

description of the local statistical correlations among the alloying atoms. Since the results of this sixteen-bond cluster have important implications to many properties, the results warrant a detailed discussion, which will be done in the next section.

4.9 Sixteen-bond Microclusters

As mentioned earlier, the 16-bond clusters shown in Figure 4.8.1 are the smallest non-overlapping clusters that are useful to examine local statistical correlations of alloy atoms in a pseudobinary alloy $A_{1-x}B_xC$. Each cluster contains four alloy atoms A and B on the vertices of a tetrahedron. The 16 bonds in a cluster include the four bonds connecting the four alloy atoms to the central C atom, and the twelve bonds connecting the four alloy atoms to twelve C atoms on the periphery. In the absence of external stresses, there are five different kinds of clusters distinguished by distinct cluster energies. These clusters can be labeled according to the numbers of A and B atoms in the cluster, namely A_4B_0 , A_3B_1 , A_2B_2 , A_1B_3 , and A_0B_4 , to be labeled by $j = 0$ to J with $J = n = 4$, following the notation in Section 4.7. Each cluster now occupies one cube in the crystal. Thus the ratio of the number of clusters M to the number of unit cells (or C atoms) N is $M/N = 1/4$. Since each cluster shares no alloy atoms with any other cluster, the total number of alloy configurations Φ in Eq. (4.7.5) for a given set of numbers of clusters (M_j) is exact. Another consequence of the fact that $n^Q = 1$, as can be seen from Eq. 4.7.16, is that some entropy terms in the free energy vanish. Any error lies in approximating the Hamiltonian as a sum of independent cluster contributions.

The effective cluster energies can be calculated in a manner similar to what was done for the one-atom clusters in Sec. 4.8. To circumvent the complicated atomic relaxation that can occur in the medium outside a cluster, the effective spring model results deduced in the single impurity case can be extended to the present case. To be more explicit, we start with a VCA medium and replace sixteen bonds in this medium by a cluster. The atoms in the cluster including the twelve peripheral C atoms are allowed to relax, but the atoms in the medium are fixed at their VCA sites. To compensate the imposed rigidity, the bond-angle restoring force

β in VFF is taken to be zero for all interactions involving those bonds connecting the peripheral C atoms to the medium. With this simplification the strain energies and atomic positions can be treated without further approximations in VFF through a minimization of the strain energy. These strain energies, according to Eq. (4.8.10), can be assigned as the strain contribution to the cluster energy.

Contributions to the cluster energies from modifications of the chemical bonds in the alloy can be estimated using the bond-orbital model (BOM) of Sec. 2.5 by calculating the changes in the metallization energy of each bond, i.e., $\Delta\epsilon_b$ in Eq. (2.5.7), from its pure crystal value (Chen and Sher, 1985). The contributions to $\Delta\epsilon_b$ due to antibonding states $|a'\rangle$ coming from the same cluster can be treated exactly, while those from the other clusters are approximated as statistically averaged values in a manner indicated in Eq. (4.8.3).

Figure 4.9.1 shows the calculated cluster excess energies as a function of x in $\text{Ga}_{1-x}\text{In}_x\text{As}$. For a small x value, the ϵ_j energies are larger for larger j , because a larger j means there are more InAs bonds in the cluster, and therefore, it is harder to fit it into a medium with a lattice constant close to that of a GaAs crystal. Similarly, at $x = 1/2$, ϵ_2 for the A_2B_2 cluster is the smallest energy because this cluster has the best match to the 50-50 VCA alloy lattice. Clearly, these energies are dominated by strain. However, it is often misleading to judge the local correlation from these energies. To do this it is better to look at the reduced cluster energies Δ_j defined in Eq. (4.7.4) because it is these energies that drive the cluster populations, given by Eq. (4.7.12).

Figure 4.9.2 shows the reduced cluster energies Δ_j for $\text{Ga}_{1-x}\text{In}_x\text{As}$. By definition Δ_0 and Δ_4 are zero. The other Δ_j 's differ drastically from the ϵ_j set, not only in magnitude but also in their x dependence. These reduced cluster energies are all negative and have a weak linear dependence on x . The three lines are nearly parallel to each other indicating the same x dependence. Δ_1 and Δ_3 are nearly equal and are less negative than Δ_2 . The negative values of the Δ_j qualitatively imply that A and B atoms do not like to segregate. These general features of cluster energies are found to be common in all III-V and II-IV alloys with appreciable (few percent)

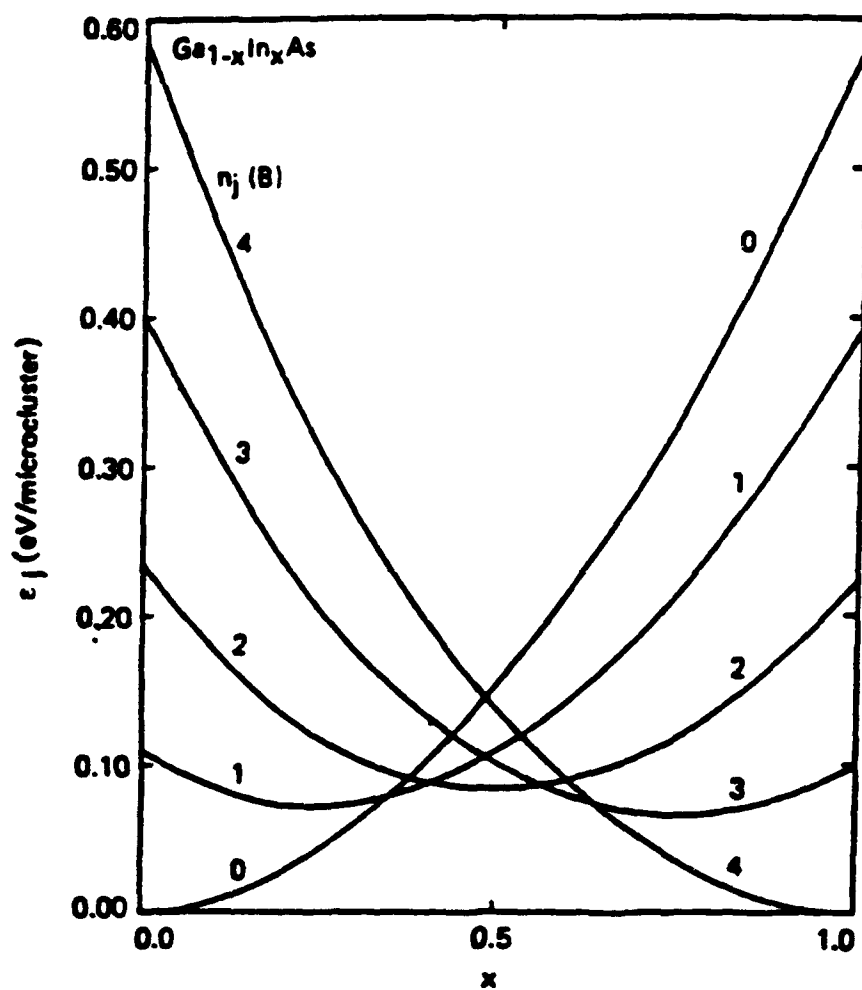


Figure 4.9.1 Cluster excess energies of the 16-bond clusters described in the text in $\text{Ga}_{1-x}\text{In}_x\text{As}$ for clusters with different numbers of In atoms.

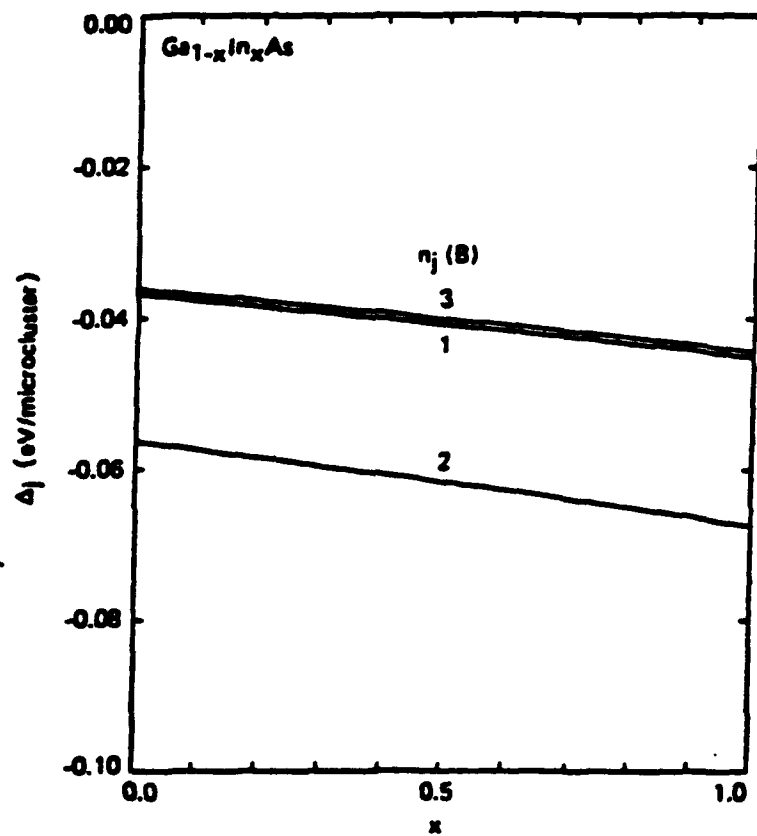


Figure 4.9.2 The corresponding reduced cluster energies Δ_j for $\text{Ga}_{1-x}\text{In}_x\text{As}$.

lattice mismatches. It is evident from Fig. 4.9.2 that δ_1 and δ_2 in Eq. (4.7.28) have opposite signs and are roughly given by $\delta_1 = -n\delta_2 > 0$, so $\Delta_j \cong -\delta_2 j(4-j) \cong \Delta_2 j(4-j)/4$. Then the energy contribution to M_B in Eq. (4.7.31) from the state of order is $n\ell (\delta_1 + nx\delta_2) \cong \Delta_2(1-x)$. When x is zero there is a small negative enthalpy contribution to the chemical potential, i.e., to the free energy when a B atom is added, but when x is unity the state of order makes no enthalpy contribution to the free energy when an extra B atom is added. For the lattice-matched alloys such as $\text{Ga}_{1-x}\text{Al}_x\text{As}$ and $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, all the energies are small and the effects caused by $\{\Delta_j\}$ are negligible.

The best way to measure the local correlation is to look at the average values of the cluster population $x_j = M_j/M$ against those in a random alloy x_j^0 given in q. (4.7.6) with g_j given in Eq. (4.7.15). The average cluster populations can be calculated from Eq. (4.7.12) after η is determined by solving Eq. (4.7.14). Figure 4.9.3 shows a plot of M_R determined from $\eta = e^{M_{BR}/kT}$ for several temperatures as a function of x . These curves were determined numerically and have none of the approximations inherent in Eq. (4.7.28). The shapes of the curves are just what would be expected from the trends predicted by Eq. (4.7.28). The reduced chemical potential for B atoms M_{BR} is large and negative for small x and large and positive for large x because of the entropy terms, and small near the center of the x distribution where the Δ_j 's are influential. Figure 4.9.4a and b show plots of the deviations from randomness, $x_j - x_j^0$ for different clusters as a function of alloy concentration x at equilibration temperatures of 600K and 1500K respectively. These two sets of curves are similar in shape but different in magnitude. As expected from the fact that $\Delta_1, \Delta_2, \Delta_3, < 0$ and $\Delta_0 = \Delta_4 = 0$, the $j = 1, 2, 3$, cluster populations, where they are large, are seen to be enhanced by the Δ_j , while the $j = 0$, and 4 populations are reduced. For reference the random populations $\{x_j^0\}$ as a function of concentration are in Fig. 4.9.4c. These cluster populations are seen to be sensitive to temperature. For example, at $x = 0.54$, the deviation decreases from 25% to 5.5% at T increases from 600 to 1500 K. For a given sample, the actual temperature to be used to calculate the cluster population depends on the "last" temperature T_f at which the alloy atoms

cease to diffuse. If T_f is larger than the sample temperature T , then the alloy is in a metastable state, and T_f instead of T should be used.

After the cluster distribution is calculated, the alloy mixing enthalpy is determined from Eq. (4.7.3). Figure 4.9.5 shows ΔE as a function of x for $\text{Ga}_{1-x}\text{In}_x\text{As}$ at four temperatures; $T = 300, 600, 1000$, and 1500 K. The ΔE is seen to be dominated by the T -independent part given by the first two terms on the right side of Eq. (4.7.3), which can be attributed to the strain energy. The T -dependence is seen to be very weak, because the Δ_j s are small. ΔE increases with T because that promotes increases of population of clusters with higher energies. The curves in Fig. 4.9.5 indicate that the approximation $\Delta E = x(1-x)\Omega$ with a constant positive Ω , such as that used in Sec. 4.7, should not introduce much error in the free energy calculation. However, a positive value of Ω does not necessarily imply

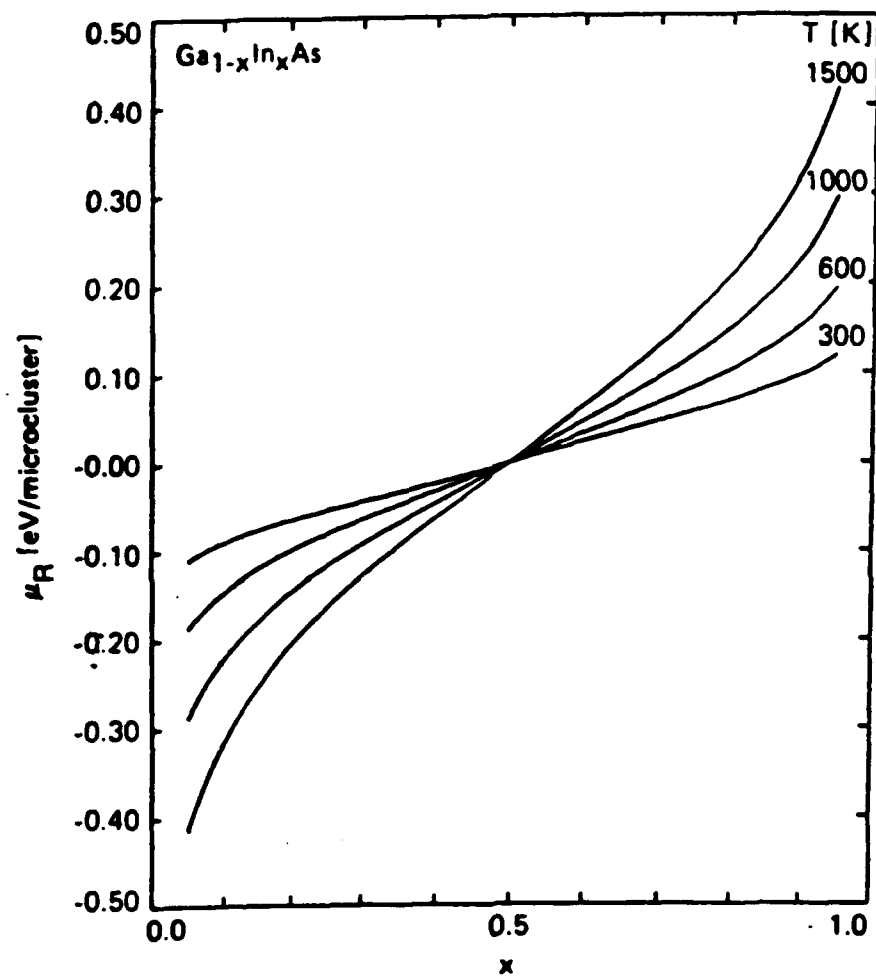


Figure 4.9.3 Reduced chemical potential μ_R as a function of concentration x for several temperatures $T = 300, 600, 1500$ K in the alloy $\text{Ga}_{1-x}\text{In}_x\text{As}$.

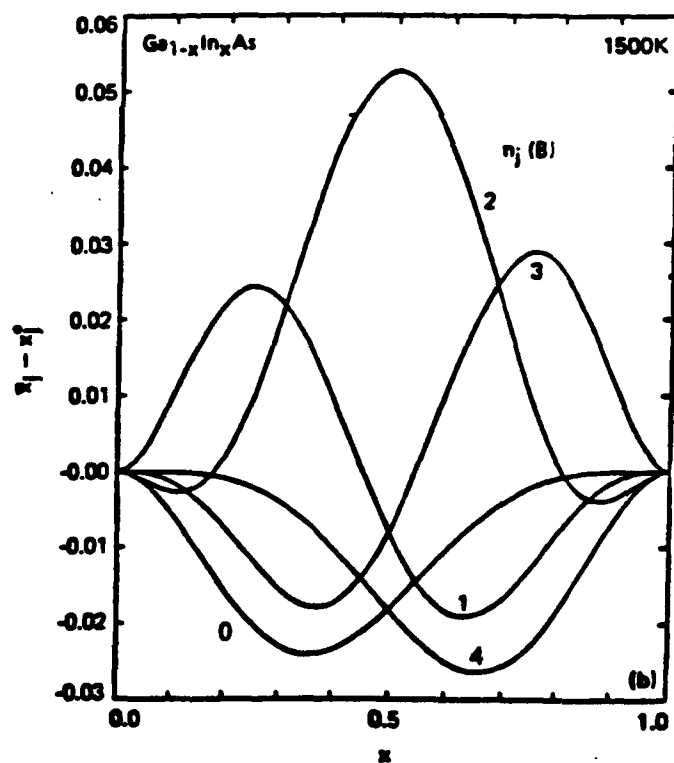
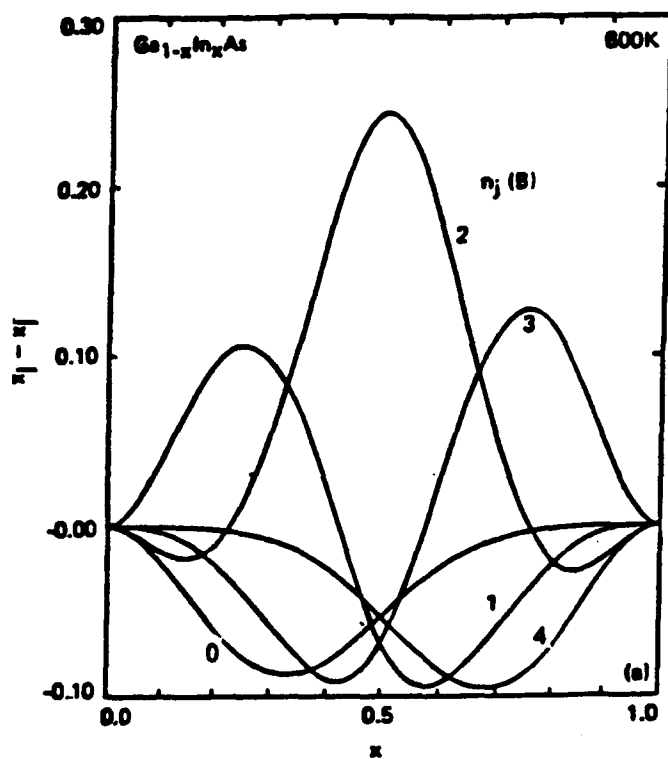
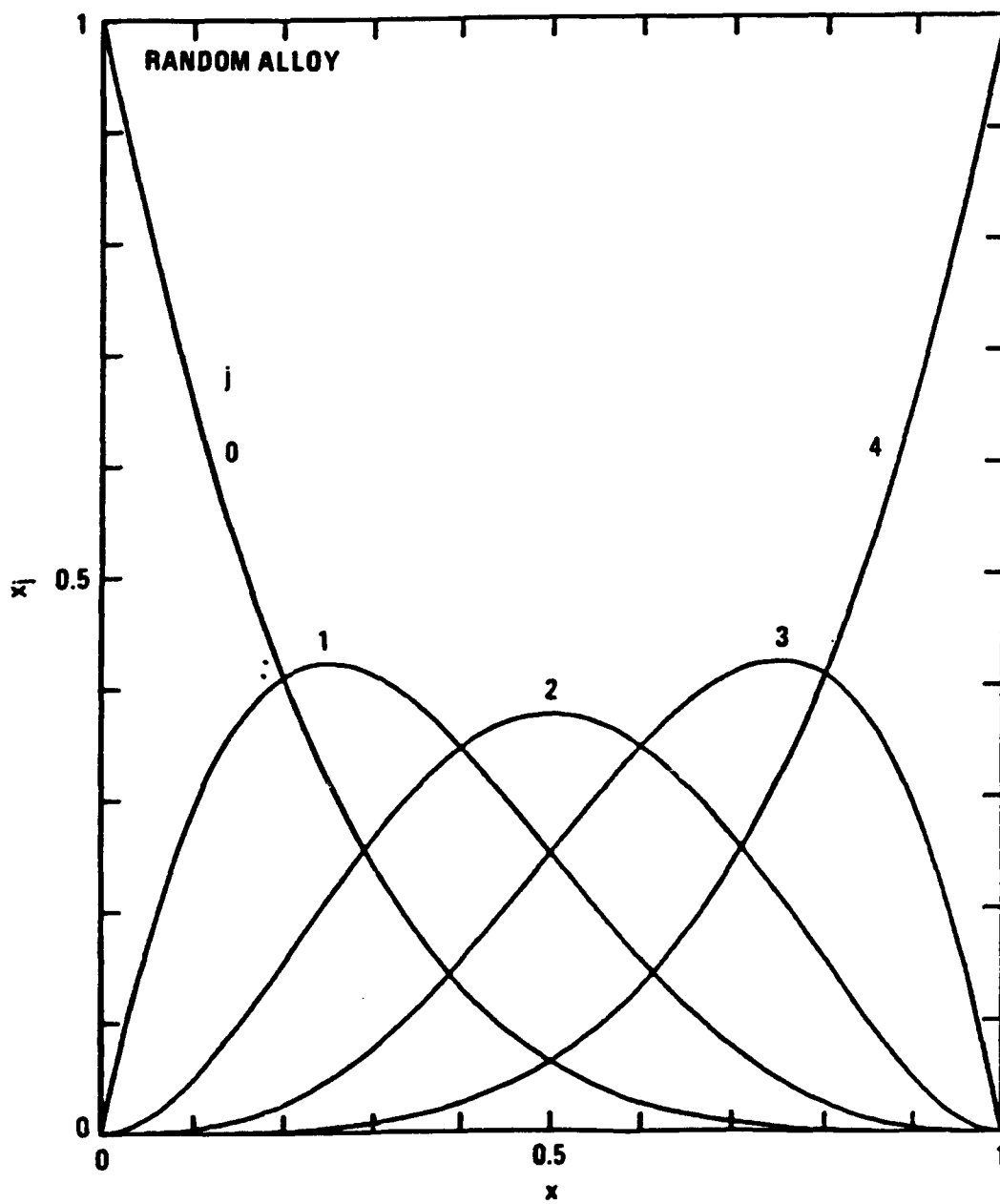


Figure 4.9.4 Deviations of cluster populations in $\text{Ga}_{1-x}\text{In}_x\text{As}$ from random distribution at (a) $T = 600 \text{ K}$ and (b) 1500 K . For reference the random populations $\{x_j\}$ are plotted against x in (c).

fig. 4.9.4c



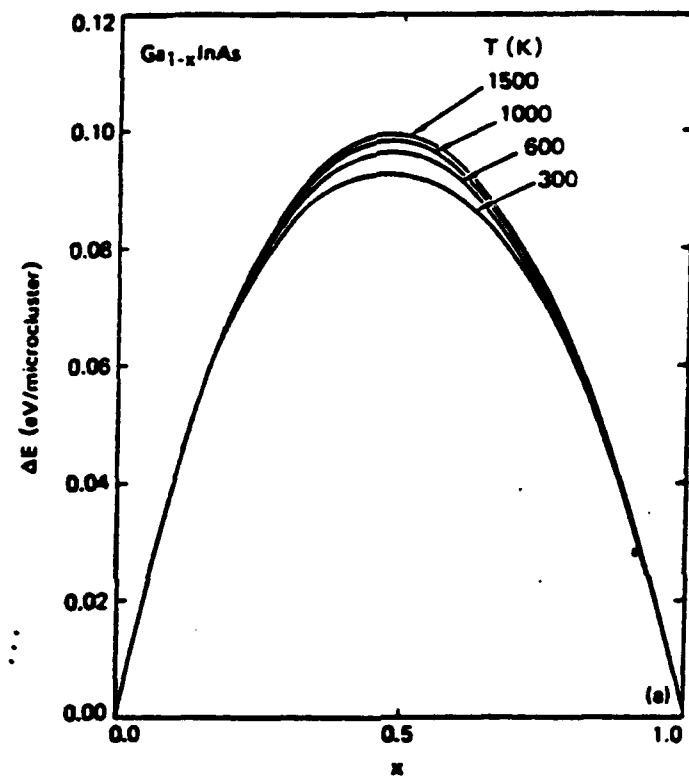


Figure 4.9.5 Mixing enthalpies of $\text{Ga}_{1-x}\text{In}_x\text{As}$ as a function of x at four temperatures.

that A and B atoms repel each other microscopically. This is evident from Figure 4.9.5 which shows enhanced population for the A_2B_2 clusters relative to those of a random alloy.

The total mixing free energy, ΔF , can then be calculated from $\Delta E - T\Delta S$ with ΔS given by Eq. (4.7.7). The results for $Ga_{1-x}In_xAs$ are shown in Figure 4.9.6. These curves show that this alloy has a miscibility gap at low temperature. These results are similar to those based on pair-potential QCA. Therefore, it is not surprising that replacement of the pair-QCA ΔF for pseudobinary solid solutions by the present GQCA results produce nearly the same phase diagrams as those shown in Fig. 4.5.4 (Patrick et al., 1987). This also shows that the phase diagrams are in the class of phenomena that are rather insensitive to details of the short range order statistical distributions. Effects sensitive to their local environment, e.g., vacancy formation, are more sensitive to the short range order state, than those like phase diagrams that are related to average properties.

A byproduct of these calculations is the bond length distribution in alloys. Figure 4.9.7a shows the bond lengths from the central C atoms to the four alloy atoms as a function of x for all five clusters considered. The spreads in the lengths of the same kind of bond in different clusters are found to be small. Figure 4.9.7b shows their average values, and the rms widths about the mean of Ga-As and In-As bonds over the cluster populations. The calculated result clearly supports the bimodal bond length distribution revealed by EXAFS.

Besides $Ga_{1-x}In_xAs$, the above studies have been carried out for several other pseudobinary alloys, including $Ga_{1-x}In_xP$, $GaAs_{1-x}Sb_x$, $Hg_{1-x}Cd_xTe$, $Hg_{1-x}Zn_xTe$, $Cd_{1-x}Zn_xTe$ and $ZnSe_{1-x}Te_x$ (Patrick et al., 1987 and 1988). Among quantities that can be checked accurately against experiment, the bond lengths and the phase diagrams from the 16-bond cluster, generalized QCA calculations compare favorably with experiments. As further evidence, Fig. 4.9.8 shows the liquidus-solidus curves for three II-VI pseudobinary alloys based on the 16-bond GQCA calculations for the solid phases (Patrick et al., 1988). Agreement between the predictions and the experiments are very good. The effective mixing enthalpy

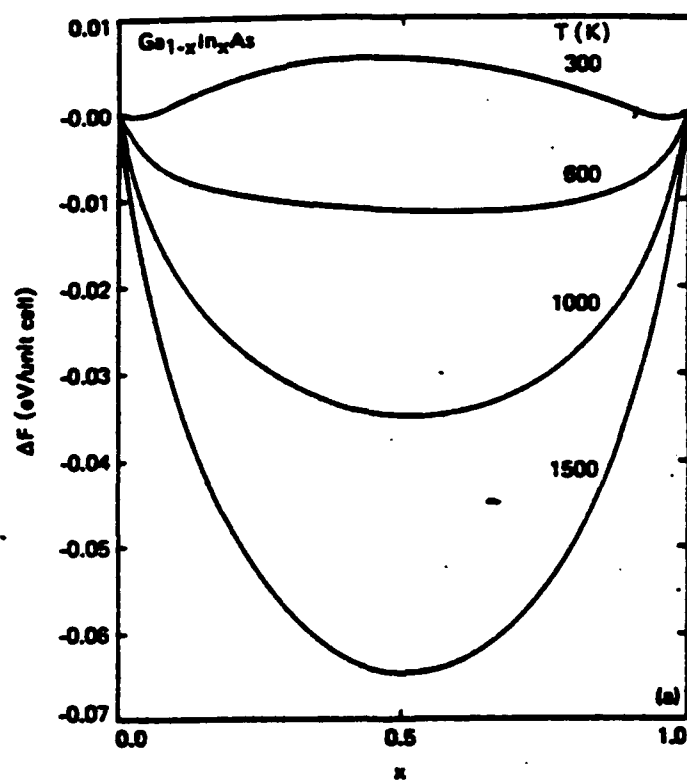


Figure 4.9.6 Mixing free energies of $\text{Ga}_{1-x}\text{In}_x\text{As}$ as a function of x at four temperatures.

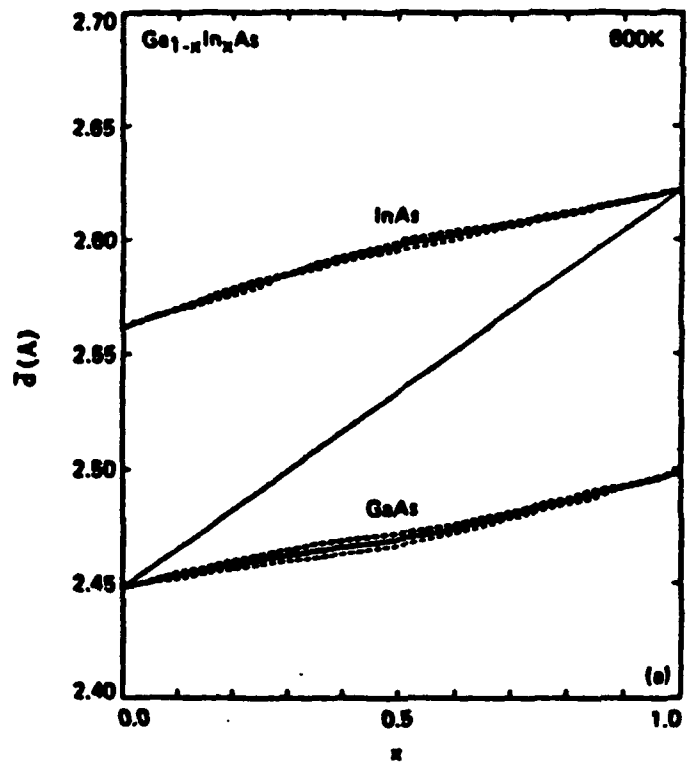
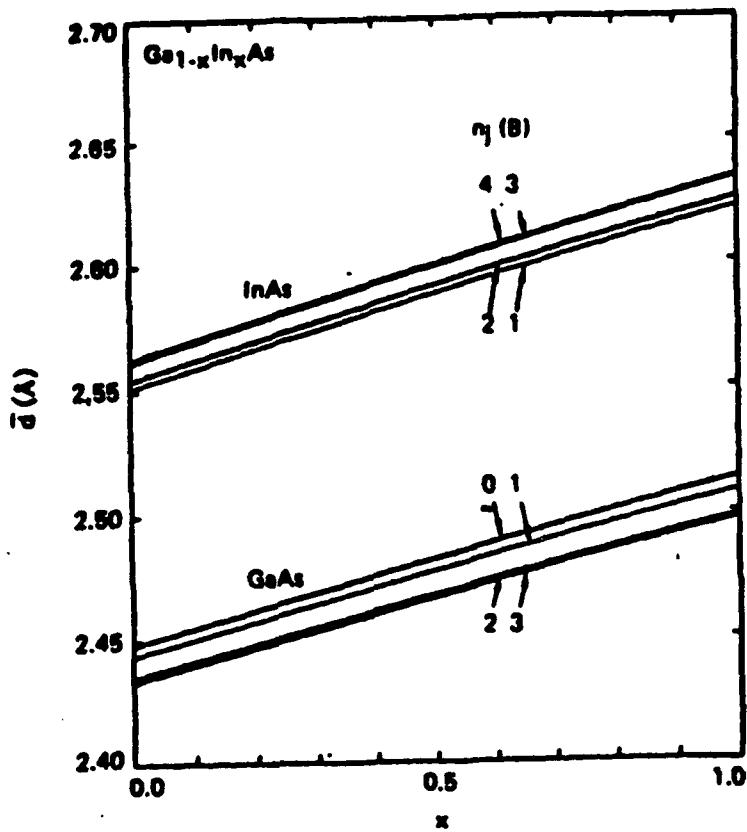


Figure 4.9.7 (a) The Ga-As and In-As bond lengths in different 16-bond clusters in $\text{Ga}_{1-x}\text{In}_x\text{As}$ as a function of x . (b) The average bond lengths at $T = 600\text{ K}$. The figure also shows the rms widths of the distribution.

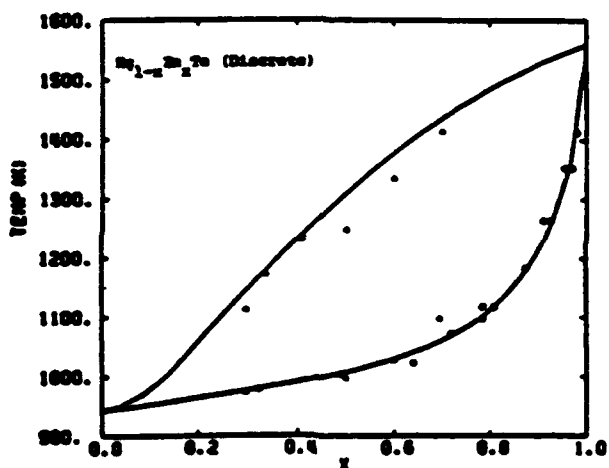
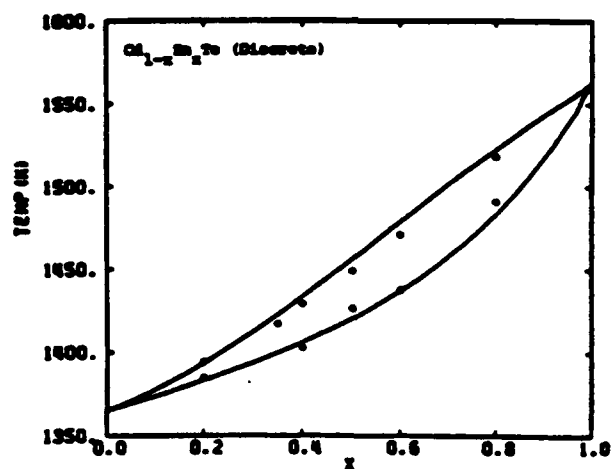
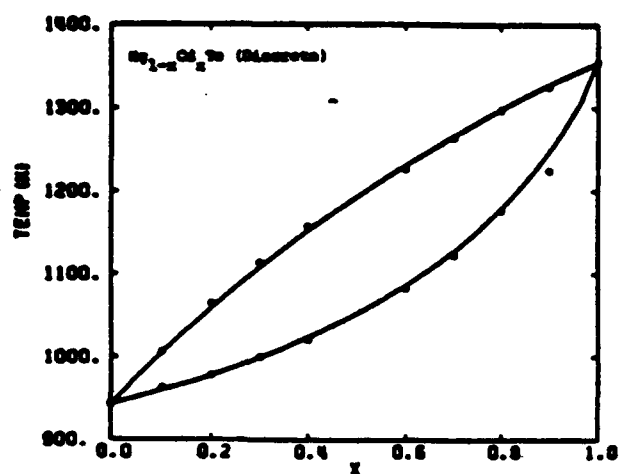


Figure 4.9.8 Calculated liquidus-solidus curves for three II-VI pseudobinary alloys using the 16-bond GQCA results for the solid phase and their comparison with experiments. After Patrick et al. (1988).

parameters Ω_{16} for completely disordered alloys, i.e. random alloys, derived from the present 16-bond model for a number of systems are listed in Table 4.6.2. These are seen to be in accord with the experimental values and with those from previous estimates. However, the attractive interactions between the alloy atoms, and the associated changes in the cluster populations found from the calculations, have not been clearly demonstrated experimentally.

4.10 Cluster Variational Method

The mixing energy models and statistical methods discussed so far have provided an adequate description of local correlation and phase diagrams in pseudobinary semiconductor alloys. All the results in these models were obtained without relying on heavy computations. However, modern computers do enable sophisticated calculations, and more rigorous calculations should be considered. This section will discuss an improved statistical theory, and the next section will review recent efforts to combine this statistical method with first-principles LDA calculations for semiconductor pseudobinary alloys.

A statistical method first proposed by Kikuchi (1951) which improves GQCA by incorporating some statistical correlation between clusters, is referred to as the cluster variational method (CVM). Like GQCA, CVM divides a crystal into clusters which are chosen to contain all the important interactions. There are several approaches to obtain the CVM cluster configurations. The method we shall consider is perhaps the simplest one. Unless specified, we shall ignore the spectator C atoms in the pseudobinary $A_{1-x}B_xC$ alloy, then the statistical counting is the same as that in the binary $A_{1-x}B_x$. We will use a familiar notation. There are N lattice sites equal to the number of atoms $N = N_A + N_B$. M is the number of basic clusters in the crystal. For example, if the clusters are pairs, then $M = zN/2$, where z is the coordination number. If the basic cluster is a square, then $M = N$ in a square lattice, but $M = 3N$ in a simple cubic lattice. The quantity of interest is the number of ways, Φ , of arranging the A and B atoms on the lattice for a specified set of the numbers of different types

of clusters $\{M_j\}$. If these clusters were all independent, then the answer would be

$$g(M,s) = M! / \left[\prod_j M_j ! \right] , \quad (4.10.1)$$

where the index s labels the basis cluster being used, e.g. pair, triangle, square, etc. However, these basis clusters in a crystal share subclusters, e.g., squares may share pairs. Therefore, the counting $g(M,s)$ has to be modified. The subclusters can also share further subclusters, and further corrections have to be made until reaching the final smallest subclusters - the single-atom clusters to be referred to as points. The ways systematic corrections are made can be best understood by going through several examples.

Consider the simplest case in which the basis clusters are pairs. Then $M = zN/2$, and the uncorrected Φ is given by $g(M,\text{pair})$ in Eq. (4.10.1). These pairs share points. The M independent pairs would contain $2M$ points, whereas there are only N points in the crystal. A logical way to correct $g(M,\text{pair})$ is to normalize it with ratio of point arrangements as follows

$$\Phi = g(M,\text{pair}) [g(N,\text{point}) / g(2M,\text{point})], \quad (4.10.2)$$

where $g(N,\text{point})$ following the definition of Eq. (4.10.1) is the number of random configurations Φ_0 in Eq. (4.2.1)*. Using the Stirling approximation $\ln N! = N \ln N - N$, the mixing entropy $\Delta S = k \ln \Phi$ becomes

$$\Delta S = -Nk \sum_j c_j \ln c_j - Mk [\sum_j h_j y_j \ln y_j - 2 \sum_j c_j \ln c_j] , \quad (4.10.3)$$

where for latter convenience the notation c_1 and c_2 stand for fractions x and $(1-x)$ respectively, $y_j = M_j/M$ is the pair probability of cluster type j , and h_j is the degeneracy of cluster type j . Note that in the above notation, M_j is the number of clusters of the type identified by specifying the arrangement of A and B atoms in the cluster, it does not include other types with the same cluster energy. Thus $h_j y_j$ is equivalent to the cluster

*In terms of the logic used to get Eq. (4.7.5), Φ equals the number of arrangements of the N_A and N_B atoms on the N sites, $g(N,\text{point})$ times the fraction of those arrangements $g(M,\text{pair})/g(2M, \text{point})$ corresponding to the pair arrangements.

population x_j in QCA. With this understanding, then the CVM is exactly the same as QCA in the pair approximation.

Next consider square clusters on a square lattice. Let M now be the total number of squares in the crystal, then $M = N$. The number of pairs in the crystal is actually $L = 2N$. If all the M squares were independent, e.g., separated from one another, then the total number of pairs would be $4M$ and the total number of atoms (points) would also be $4M$. The first correction to $g(M, \text{square})$ comes from the following pair normalization factor

$$f_{\text{pair}} = \frac{g(L, \text{pair})}{g(4M, \text{pair})} . \quad (4.10.4)$$

However, both the denominator and numerator of this factor have to be further corrected with point normalizations in a form similar to that inside the bracket of Eq. (4.10.2). Explicitly, the corrected Φ takes the form

$$\Phi = g(M, \text{square}) \left(\frac{g(L, \text{pair})}{g(4M, \text{pair})} \right) \left(\frac{g(8M, \text{point})}{g(4M, \text{point})} \right) / \left(\frac{g(2L, \text{point})}{g(N, \text{point})} \right) . \quad (4.10.5)$$

In this equation the first bracket is the pair normalization, the second bracket is the point normalization to correct $g(4M, \text{pair})$, and the last bracket is to correct $g(L, \text{pair})$. The associated mixing entropy becomes

$$\begin{aligned} \Delta S &= -k \left[(8M - 4M - 2L + N) \sum c_j \ln c_j + (L - 4M) \sum h_j y_j \ln y_j + M \sum g_j z_j \ln z_j \right] \\ &= -kN \left(\sum c_j \ln c_j - 2 \sum h_j y_j \ln y_j + \sum g_j z_j \ln z_j \right) , \end{aligned} \quad (4.10.6)$$

where y_j and z_j are respectively the pair and square probabilities, and h_j and g_j are their respective degeneracies.

Now it becomes straightforward to write down the expression for ΔS for square clusters in a (3-dimensional) simple cubic lattice, for which $M = 3N$, $L = 3N$, and the expression for Φ in Eq. (4.10.5) is still valid, so

$$\Delta S = -kN \left(7 \sum c_j \ln c_j - 9 \sum h_j y_j \ln y_j + 3 \sum g_j z_j \ln z_j \right) . \quad (4.10.7)$$

As a final example, consider the tetrahedron clusters in a fcc lattice. There are a total of $M = 2N$ tetrahedral clusters, and $L = 6N$ pairs in the crystal, and there are 6 sides and 4 atoms in a tetrahedron. Eq. (4.10.5) has to be modified slightly for the present case to read

$$\Phi = g(M, \text{square}) \left(\frac{g(L, \text{pair})}{g(6M, \text{pair})} \right) \left(\frac{g(12M, \text{point})}{g(4M, \text{point})} \right) / \left(\frac{g(2L, \text{point})}{g(N, \text{point})} \right) . \quad (4.10.8)$$

The entropy is then given by

$$\Delta S = -kN (5\sum c_j \ln c_j - 6\sum h_j y_j \ln y_j + 2\sum g_j z_j \ln z_j) . \quad (4.10.9)$$

Up to now we did not have to distinguish between binary and pseudobinary alloys. However, for the tetrahedron case there are two different kinds of tetrahedra in pseudobinaries, one contains a central C atom, and the other is empty. If the energies are assumed to be associated only with bonds, then we can ignore those tetrahedra without a central C atom. Then $M = N$ should be used in Eq. (4.10.8) to give

$$\Delta S = -kN (-3\sum c_j \ln c_j + \sum g_j z_j \ln z_j) . \quad (4.10.10)$$

Note that the pair contribution disappears from this expression. This is understandable, because the clusters treated this way share only points. It is important to point out that Eq. (4.10.10) is exactly the same as the GQCA result in Eq. (4.3.7) for the case of tetrahedra with the identification of x_j with $g_j z_j$. This example demonstrates that alloy statistics for pseudobinaries are not necessarily identical to that for binaries.

It is useful to observe a rule that governs the coefficients in front of the summation signs inside the brackets for all CVM expressions for ΔS : If we multiply each coefficient by the number of atoms in the cluster that term represents, and sum up these products, the result is one. This rule is useful for checking if the configurations are properly normalized.

The equilibrium values of cluster probabilities such as $\{y_j\}$ and $\{z_j\}$ are determined by minimization of the mixing free energies in a manner similar to GQCA. Likewise, not all cluster probabilities are independent. Since there are more variables to be determined and more constraint

conditions to be satisfied in CVM than in GQCA. Kikuchi (1981) worked with the grand potential $\phi = F - \sum \mu_j N_j$, and designed an iterative procedure, the so-called "natural iteration method", to attack the problem. Impressive results have been obtained from CVM with small clusters. For example, Table 4.2.1 shows that CVM predicts critical temperatures very well for the square lattice using square clusters, and for the fcc lattice using tetrahedron clusters. There are also examples where CVM obtains correct phase diagrams, whereas the Bragg-Williams approximation fails quantitatively (Kikuchi, 1974). This method has recently been incorporated in the phase diagrams studies of semiconductor alloys using energies derived from *ab-initio* LDA calculations. The results will be discussed in the next section.

4.11 Ab-Initio Calculations

Because self-consistent LDA has been successful predicting the structural properties of ordered semiconductors, it is desirable to attempt to apply this *ab-initio* theory to alloys. There are several challenges in such an undertaking. First, the mixing enthalpies in semiconductor alloys are typically of order of 1 Kcal per mole or 1/23 eV per atom, but the total energies i.e., each term on the right side of Eq. (4.7.2) are many orders larger. Very precise and accurate numerical calculations are required to get meaningful results. However, if carried out precisely, LDA has been demonstrated to be able to predict changes in energies correctly, despite its larger error in predicting total energies. The next, and more challenging difficulty, is alloy disorder. To compute the total energy of an alloy quantum mechanically one needs to solve the Schrodinger equation for a Hamiltonian which does not have lattice translational symmetry, which is so indispensable in the traditional band-structure theory. Finally, to use a statistical theory such as CVM, the excess energy has to be composed of short-ranged multi-site correlation energies, including only the single-site, pair, etc., up to a manageable cluster size containing only a handful of sites. To circumvent these difficulties, Connolly and Williams (1983), working on metal alloys, proposed that these multi-site correlation energies be deduced from ordered systems that are composed of the same atoms. This scheme allows a direct application of the first

principles theory in the calculation of the energy parameters. This theory has been extended to semiconductor alloys, including considerable refinements, by Wei et al. (1990) and Ferreira et al. (1989) (to be referred to as WF), and respectable results have been obtained for the phase diagrams and alloy equilibrium properties. Their main results are summarized below.

The cluster excess energies in WF are calculated as a function of relative cell volume. Figure 4.11.1 shows an example for $\text{Ga}_4\text{As}_n\text{Sb}_{4-n}$ clusters. These energy curves are similar to the x -dependent excess energies shown in Fig. 4.9.1, because the alloy cell volume was scaled linearly with x . The overall implications to these excess energies are also very much in accord with what we learned from Sec. 4.9. The magnitudes of mixing enthalpies are generally dominated by terms driven by lattice constant mismatches. For the lattice-matched alloys, the excess energies are very small, their effects fall within calculational uncertainties, and therefore may be neglected. For lattice mismatched systems, ΔE is positive and as we have seen in Sec. 4.9 is dominated by a temperature independent strain contribution. The reduced excess energies that govern the atomic statistical correlation are small and negative. This implies anticlustering, as shown in Figure 4.11.2 for the deviation of cluster populations from random-alloy values. These curves are qualitatively the same as those obtained in Sec. 4.9. The mixing enthalpy parameters Ω for random alloys (at $T = \infty$) were calculated and are listed in Table 4.6.2, where they can be compared with previous estimates. The miscibility gaps and spinodals are also calculated and are shown in Figure 4.11.3. These curves display considerable asymmetry about $x = 1/2$, a behavior also found from the experimental data for $\text{GaSb}_{1-x}\text{As}_x$. The calculated critical temperatures, indicated as T_{MG} , and the concentration x_{MG} where the bimodal curves peak, are tabulated in Table 4.11.1. For systems for which experimental data are available, the agreement between theory and experiments are good. Finally, the predication of the bimodal bond length

fig. 4.11.1

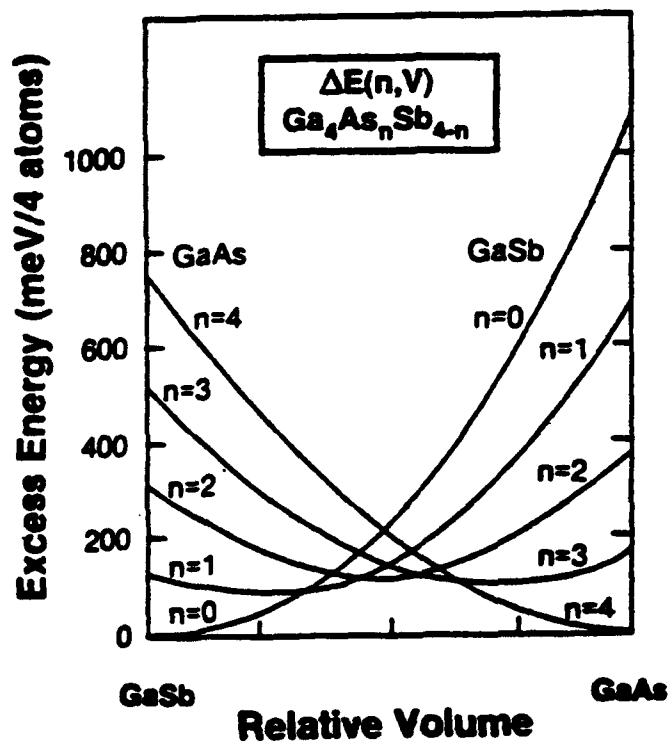


Figure 4.11.1 Volume dependent excess energies for $\text{Ga}_4\text{As}_n\text{Sb}_{4-n}$ clusters obtained by Ferreira et al. (1989).

fig. 4.11.2

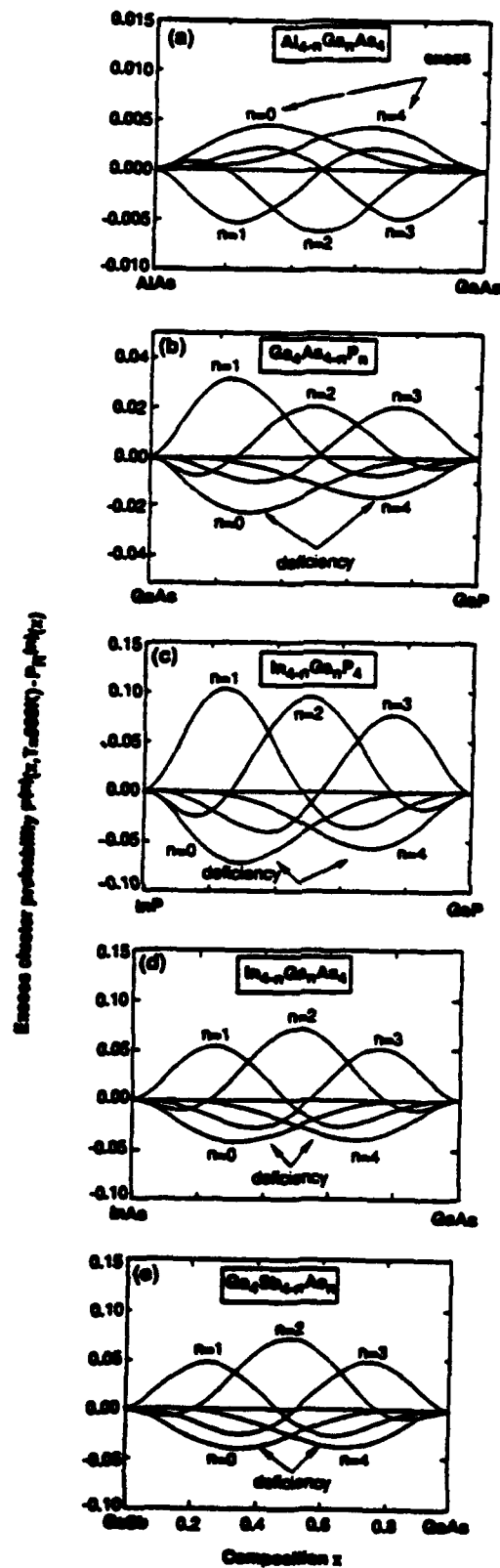


Figure 4. 11.2 Deviations of cluster populations from random distributions for five III-V pseudobinary alloys calculated by Wei et al. (1990).

fig. 4.11.3

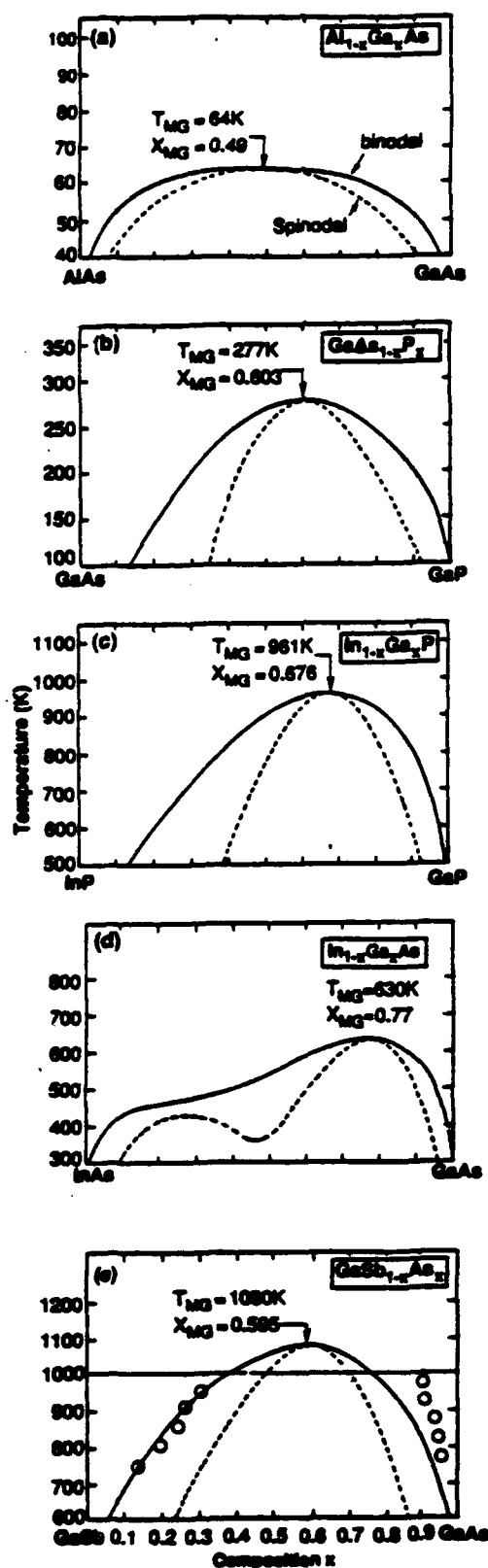


Figure 4.11.3 Miscibility gap and spinodal curves for five III-V pseudobinary alloys calculated by Wei et al. (1990).

distributions are similar to those predicated from the 16-bond cluster model in Sec. 4.9. However, we should note that in order to produce the kind of accuracy in the energies required for the phase diagram calculation very detailed and careful computational efforts are needed. The results presented here represent the state of the art in the LDA calculations.

Table 4.11.1

Calculated critical temperatures T_{MG} for the binodal and spinodal curves and the corresponding compositions x_{MG} compared to experiments.

System	Theory		Experiment	
	$T_{MG}(K)$	x_{MG}	$T_{MG}(K)$	x_{MG}
$Al_{1-x}Ga_xAs$	64	0.49		
$In_{1-x}Ga_xP$	961	0.676	933	0.62
$In_{1-x}Ga_xAs$	630	0.77		
$GaAs_{1-x}P_x$	277	0.603		
$GaSb_{1-x}As_x$	1080	0.595		
$Hg_{1-x}Cd_xTe$	84	0.60		
$Hg_{1-x}Zn_x$	455	0.56		
$Cd_{1-x}Zn_x$	605	0.623		

After Wei et al. (1990)

However, the Connolly-Williams approach is only a kind of interpolation between *ab-initio* energies of the ordered systems and alloys. The multi-site correlation energies deduced from this approach represent interpolation parameters for the total energies. They depend on the ordered systems chosen in the parametrization. The volume dependent cluster energies are unlikely to be sufficient for calculating alloy

properties sensitive to the cluster shapes, e.g., shear moduli and lattice vibrations. These are fine points that require improvements. A different way to approach the *ab-initio* calculation is to attack the disordered problem directly. If the fluctuations of the alloy potentials from their virtual crystal approximation (VCA) is small, then the next leading correction to VCA can be obtained from perturbation theory. This should work for most semiconductor alloys except for systems with large potential fluctuations such as $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (Spicer et al., 1982). A more general but more difficult approach is to extend an effective modern alloy theory, the coherent-potential approximation (CPA) (to be discussed in Chapter 5), to clusters and to achieve triple self-consistency: consistency between the cluster distribution and the Hamiltonian, between the Hamiltonian and the electron density, and between the self-energy operator Σ in the cluster CPA theory and the potential fluctuations. Currently this theory has been carried out only for metal alloys, and only within the single-site KKR-CPA (Gyorffy and Stocks, 1979) with a random distribution. There is also a theory developed along with CPA, the so-called generalized perturbation method GPM (Ducastelle and Gautier, 1976), which is intended to systematically calculate multi-site correlation energies from the CPA results. However, to use GPM in semiconductors, it needs to be extended to include total energy contributions from other than the band structure term in EQ. (2.2.13) and to allow lattice distortions. Major work is needed if this approach is to achieve the same degree of rigor for disordered alloys as self-consistent density functional theory now enjoys dealing with crystalline semiconductors.

Appendix G

**HgCdTe STATUS REVIEW WITH EMPHASIS ON CORRELATIONS,
NATIVE DEFECTS, AND DIFFUSION**

HgCdTe status review with emphasis on correlations, native defects and diffusion†

A Sher‡, M A Berding‡, M van Schilfgaarde‡ and An-Ban Chen§

‡ SRI International, Menlo Park, CA, 94025 USA

§ Auburn University, Auburn, AL, 36849 USA

Abstract. We review the current status of knowledge of fundamental properties of the alloy $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$. The most vexing questions are about its correlation state. Several different experiments now suggest it is highly correlated, but no theory predicts this result. We also discuss other properties, including dislocations at interfaces, the residual donor, worms, surface segregation and its impact on passivation, and concentration fluctuations. The forces driving these phenomena, where they are known, will be presented. Most of the paper focuses on the following: correlations; native defects, formation enthalpies and entropies; native defect equilibria with mercury gas and with tellurium inclusions; and self-diffusion coefficient activation energies including its contribution from migration energies. We will take advantage of new first-principles, high-accuracy calculations to help explain the experimental situation. The calculations predict that the main native defects found in alloys equilibrated at low Hg pressures are Hg vacancies, while at high Hg pressures they are Hg interstitials, and, surprisingly, Hg antisites.

1. Introduction

As we have been told in this conference [1], sophisticated focal plane arrays (FPAs) of sizes as large as 256×256 elements have been demonstrated on $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ materials grown epitaxially by several means on various substrates. Control of the material and its processing is reaching a stage where practical yields of arrays with adequate performance for some applications are nearly at hand, so these array costs are becoming acceptable. While the community is now well into a manufacturing phase, there remain many unanswered questions about the nature of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloys, questions whose resolution would improve the performance, yield and stability of focal plane arrays. The purpose of this paper is to review the current status of our knowledge of this alloy system [2, 3], with the aim of suggesting potential mechanisms for the uncertain properties and additional experiments to test these hypotheses. We also will demonstrate the merit of recently developed first-principles computational methods [4] to help settle outstanding problems associated with the material which will eventually lead to accurate processing and performance models. Figure 1 depicts schematically the major issues that will be addressed. They were selected because of their potential impact on performance of FPAs. The figure shows a passivated HgCdTe sample on a substrate. On the right is a graph of a vertical composition profile through one region of the sample, and on the left is a

blow-up of a 'microcluster'. The major topics identified are point defects [2], correlations [3], 'worms' [5], misfit dislocations [6, 7], substrates, extended defects [1], inclusions [7], bulk concentration fluctuations [8] and surface segregation [9]. Other topics such as metallization [10], particularly to p-type material, could have also been included in this list, but we limit our discussion to those mentioned above. We begin by summarizing below the most important aspects of each of these topics. A selected set will be greatly expanded upon in later sections of the paper.

1.1. Point defects

Most $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloys are Te rich as-grown [1, 2]. The excess Te is accommodated by Hg vacancies (designated V_{Hg}) or as Te solid inclusions (Te_s) [7, 11]. The Hg vacancies are shallow acceptors (possibly negative-U-centre double acceptors) [12]. With the exception of some MBE or MOCVD material, as-grown material is p-type, both at room temperature and at 77 K [13, 14]. This material can usually be annealed at low temperature ($\sim 250^\circ\text{C}$) in an Hg overpressure to fill the V_{Hg} sites and to dissolve the Te inclusions [7]. When this is done the material converts to n-type with carrier concentrations that are remarkably low, from a few times 10^{14} cm^{-3} to a few times 10^{15} cm^{-3} in high-purity material [14]. The n dopant(s) (called the residual donors) are unknown. There is evidence [14] that the principal residual donor is an impurity rather than a native defect, but because it is so pervasive the evidence is not completely convincing. The residual donor concentration is unknown, uncon-

† This work was supported in part by NASA Contract NAS1-18226, ONR Contract N00014-88-C-0096, and AFOSR Contract 49620-88-K-0009.

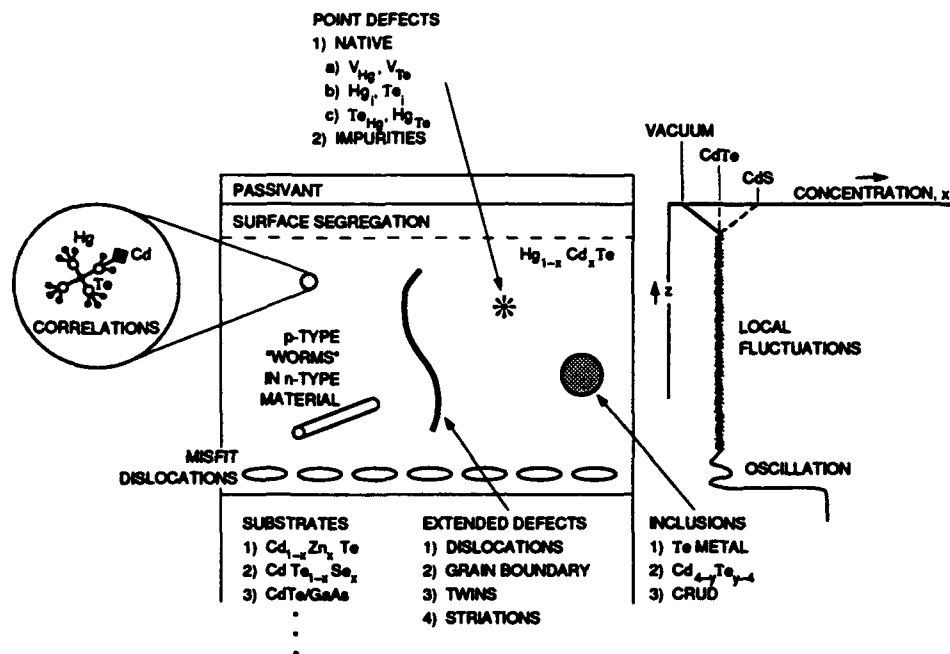


Figure 1. Schematic features of HgCdTe alloys.

trolled, and sets FPA limits for most operating modes. Identification and control of the residual donor would undoubtedly be beneficial. It has been demonstrated by diffusion measurements that samples annealed in higher Hg pressures contain Hg interstitials (Hg_i) [15, 16], but while they are likely to be donors, the electrical state of an Hg_i has never been directly identified. It has been speculated that Te interstitials (Te_i) and antisites (Te_{Hg}) [17] are active recombination centres, but direct evidence on their deep states is not available. The Hg antisite (Hg_{Te}) has not been previously suggested as a high-concentration defect, but we will present arguments in a later section that it may be present in concentrations sufficient to affect devices [2].

So far we have mentioned only native defects, but impurity doping is now being brought under control in HgCdTe alloys. It has proved to be difficult to gain control of both donor and acceptors, but methods are now in hand. Well behaved impurities are In [18] on cation sites serving as donors, and As on Te sites serving as acceptors. They are inserted into the material during epitaxial growth. Ion implantation and diffusion doping have not been successful enough to be used in any manufacturing process. Partial explanations for these observations have been published, but refinements are needed to make them quantitative. Many devices are still made utilizing V_{Hg} acceptors and residual donors [19].

1.2. Correlations

It is now well recognized that semiconductor alloys of the form $A_{1-x}B_xC$ are rarely truly random. The A and B atoms do not occupy their sublattice in the zinc blende structure at random but are somewhat correlated, so

there is a degree of short-range order, and in some cases even long-range order [20, 21]. The principal driving terms for correlations are bond length and chemical difference between the constituents AC and BC. In $Hg_{1-x}Cd_xTe$ the bond lengths of HgTe and CdTe are nearly equal, and the differences in their chemical terms are also small, so that correlations are predicted to be small. However, a high degree of correlation has been deduced using five different experimental techniques (Te nuclear magnetic resonance (NMR) [22], Raman spectroscopy [23, 24], infrared (IR) reflection spectroscopy [25] and x-ray diffraction [26]). In the concentration range near $x = 0.25$, the correlations are such that the material is tending toward an ordered structure rather than spinodal decomposition. Those findings may have important consequences for devices, in particular if the correlation state varies spatially. Such correlation fluctuations may be a source of spatial variability of material properties that occurs even if the composition is uniform, with the variation affecting the uniformity of the bandgap, native defects and impurity concentration, mobilities, etc. The magnitudes of such variations may differ greatly from one property to another. For example, one would expect relatively slight bandgap variations, but rather more substantial differences in vacancy concentrations and diffusion coefficients between regions of different correlation states. This occurs because band structures are properties that depend on site occupations averaged over many lattice spacings, but vacancy formation energies depend sensitively on the local atom arrangement around the vacant site. Both the experimental and theoretical understanding of the correlation state of HgCdTe need to be clarified. A more detailed exposition of this problem is presented in section 2.

1.3. 'Worms'

Material grown by LPE and MOVPE and annealed to n type often exhibits low carrier concentrations $\sim 10^{15} \text{ cm}^{-3}$ but with electron mobilities far below those of bulk or MBE-grown materials. If these materials are intentionally impurity doped to be slightly more n type, their mobilities return to the high values obtained by other growth methods. The currently accepted, although unconfirmed, explanation for this phenomenon is that after the Hg overpressure anneal, there are 'woodworm-like' domains in the n-type material that remain p type [5]. These regions produce p-n junctions, and mobilities measured in a Hall measurement exhibit a mixed effective mobility that appears low. Impurity doping converts these p-type worms to n type, so while there is still a spatial variation of the doping concentrations, all the material is n type so the measured Hall mobility is higher. From the viewpoint of device limits, such impurity doping has two potential deleterious effects. First the minimum carrier concentration that can be used is increased, and second there is a spatial variation of the carrier concentration.

The worms are an appealing explanation for the observed transport properties, but if the goal is to improve devices made from LPE material then we must understand why the worms form and how to eliminate them during growth, or in subsequent processing. Do the worms form because they correspond to regions of low residual donors, or to regions where the V_{Hg} formation energies are small so that they do not anneal completely? Are the worms regions where the correlation state differs from the remaining material? Why are worms more prevalent in LPE or MOVPE material than in higher-growth-temperature bulk material or lower-growth-temperature MBE material? These and other questions are still unanswered.

1.4. Misfit dislocations

Substantial misfit dislocation densities are always found at the interface between $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ and substrate materials, e.g. $\text{Cd}_{1-x}\text{Zn}_x\text{Te}$, even when the two materials are perfectly lattice matched [6, 7]. This situation also occurs to a lesser but still significant extent at heterojunctions. The reasons for these anomalous dislocation densities remain to be determined. Because dislocations have been demonstrated to have a deleterious impact on minority carrier lifetimes, their effect is of particular concern at heterojunctions where they degrade R_0A products [27-29].

1.5. Substrates

This topic has been extensively reviewed in other conference papers [1]. The ideal substrate material would be a large-area, IR-transparent insulator that is matched to the active layer by lattice constant, thermal expansion coefficient and chemical potential, and on which signal processing circuitry could be integrated. This idealization

does not exist for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, so compromises must be made. The materials CdZnTe , and GaAs or Si with buffer layers, are the current leading contenders.

1.6. Extended defects

A variety of extended defects, grain boundaries, anti-phase boundaries, twin planes, threading dislocations, striations, etc have all been directly observed in transmission electron microscope (TEM) studies or their existence deduced from indirect evidence. Their impact on device characteristics is not well established, but is certainly never helpful. Because large-area FPAs are being made, extended defects obviously are relatively dilute or ineffective in current device-grade material. However, improvements are still possible.

1.7. Inclusions

We have already mentioned Te inclusions, and they will be analysed more extensively in later sections of this paper. They always have an array of dislocations around them [7], are thought to serve as getters for some impurities, and, as we suggest later, are likely to have a non-equilibrium atmosphere of V_{Hg} and Te_{Hg} around them [2]. When they are dissolved in the Hg anneal, some of their associated dislocations climb to surfaces [7] and are eliminated, although some remain; the gettered impurities may remain as inclusions, and the larger Te₂ inclusions may not be completely dissolved. Means to avoid Te₂ inclusion formation in as-grown material are most desirable.

1.8. Bulk concentration fluctuations

Even if an alloy is random, any segment in space has a Bernoulli probability concentration distribution [30]. Thus one expects small-scale concentration fluctuations if there is nothing to suppress them. Because of the near lattice match between HgTe and CdTe, a region of space with a concentration that differs significantly from the average produces no long-range strain field and is therefore not strain suppressed. The same effect has been found in the lattice-matched alloy $\text{Al}_{1-x}\text{Ga}_x\text{As}$ [31]. However, in a lattice-mismatched material like $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$, such short-range concentration fluctuations are suppressed. One consequence of this is that in HgCdTe the exciton line is very broad, while in HgZnTe it is narrow [8]. This effect should be particularly troublesome in VLWIR material where the bandgap is very small and fluctuations could produce semi-metal shorting domains.

1.9. Surface segregation

In general, an alloy in equilibrium will not have a uniform concentration near an interface [9]. This is most easily understood at a vacuum surface where there are dangling bonds. The material can minimize its free energy by having the alloy species that makes the weakest bonds concentrate at the surface. Also in lattice-

mismatched alloys the low-concentration species will be driven to the surface by 'strain release' because they experience less strain energy at the surface than in the bulk. In the near-lattice-matched $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloy, strain release plays little part, and therefore surfaces are Hg rich, because Hg makes weaker bonds to Te than does Cd. Detailed predictions are that for $x < 0.4$ the surface layer will have concentrations in the semi-metal range returning to the bulk concentration within a few atomic layers. Thus, in equilibrium, MWIR and LWIR HgCdTe alloys will have Hg-rich semi-metal surfaces. A CdTe-passivated HgCdTe alloy will have a uniform concentration up to the interface because, both in the bulk and at the interface, cations will be surrounded by four Te atoms, so there is no driving force for segregation. There is also little driving force for segregation at a ZnTe interface with HgCdTe . These conclusions are true only if interdiffusion is prohibited across the interface.

Because there is less tendency for interdiffusion between ZnTe and $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ than between CdTe and the alloy, we have suggested that a few atomic layers of ZnTe followed by a thicker CdTe layer should be a superior passivant.

A CdS/ HgCdTe interface should be Cd rich because the CdS bond strength exceeds that of HgS by more than that between CdTe and HgTe. A similar argument applies to oxides, except that the lattice-constant mismatch becomes so large that dislocations and grain boundaries begin to exert a much larger influence on the net result.

The rapid return of the interface concentration to its bulk value (a few atomic layers) is predicted to occur only above an alloy's order-disorder transition temperatures [3, 32]. Below this temperature, long-range large-excitation concentration oscillations should occur. Because of its expected low order-disorder temperature, one would tend to discount such phenomena from consideration in HgCdTe . However, the observations of large correlation effects makes us pause. If these effects are confirmed, then there may be comparatively long-range concentration fluctuations adjacent to an epitaxial layer-substrate interface, which may help to explain the anomalous misfit dislocation densities that are observed. This suggestion (first made by Spicer *et al* [10], but without a mechanism identified) is highly speculative, and while concentration oscillations driven by surface segregation and interface boundary conditions are a natural consequence of a higher than expected order-disorder transition temperature, there is no direct evidence to support its occurrence at a $\text{HgCdTe}/\text{CdTe}$ interface. Relevant experiments would be helpful.

The remainder of this paper will be devoted to an exposition of our current understanding of correlations, native defects and their diffusion.

2. Review of correlations

We [21] and others [33-35] have demonstrated that there are always correlations in alloy semiconductors. Here we summarize our previous work, focusing atten-

tion on those aspects that may help us to identify what may be missing from the interactions driving correlations in HgCdTe . We consider an average population \bar{x}_j of five-atom anion-centred clusters including 16 bonds (most others treat four bonds only). The population of five-atom clusters of the form $\text{A}_{4-n}\text{B}_n\text{C}$ can be shown to be grand canonical ensembles of the form

$$\bar{x}_j = g_j \exp[(\mu_B n_j - \varepsilon_j)/kT] / q(\{\varepsilon_j\}, \mu_B) \quad (1)$$

where the partition function is

$$q(\{\varepsilon_j\}, \mu_B) \equiv \sum_{j=0}^J g_j \exp[(\mu_B n_j - \varepsilon_j)/kT] \quad (2)$$

and g_j is the degeneracy of cluster $j = 0, 1, \dots, J$ ($4 \leq J \leq 2^4$), n_j is the number of B atoms on cluster j , ε_j is the excess energy of cluster j relative to the virtual crystal approximation (VCA) average $\bar{\varepsilon}_{\text{VCA}} = (1-x)\varepsilon_A + x\varepsilon_B$ (called Δ_j by Sher *et al* [21]), ε_A and ε_B are cluster energies of the pure AC and BC compounds, k is the Boltzmann constant and T is the absolute temperature.

If the clusters do not have their normal counting degeneracies

$$g_j^0 = \binom{4}{n_j}$$

split, then we have $g_j = g_j^0$ the binomial coefficient, and $J = 4$. The chemical potential is set by insisting that the composition x is correct

$$4x = \sum_j n_j \bar{x}_j. \quad (3)$$

If one makes a transformation to the reduced excess-energies representation

$$\Delta_j = \varepsilon_j - \left(1 - \frac{n_j}{4}\right) \varepsilon_0 - \frac{n_j}{4} \varepsilon_j \quad (4)$$

(called Δ'_j by Sher *et al* [21]) then \bar{x}_j becomes

$$\bar{x}_j = g_j \exp[(\mu'_B n_j - \Delta_j)kT] / q(\{\Delta_j\}, \mu'_B) \quad (5)$$

where

$$\mu'_B = \mu_B - (\varepsilon_j - \varepsilon_0)/4. \quad (6)$$

Note that we have $\Delta_0 = \Delta_j = 0$, so at least two members of the set $\{\Delta_j\}$ vanish. More importantly, only the reduced excess energies actually drive the cluster populations. We shall see shortly that the energies of the $\{\Delta_j\}$ set are often five to ten times smaller than those of the $\{\varepsilon_j\}$ set. Most of the strain contribution to the $\{\varepsilon_j\}$ set subtracts out and contributes little to the \bar{x}_j values.

There is an additional effect. It can best be appreciated when stated as a theorem: If

$$\Delta_j = \Delta_0 + n_j \delta \quad \text{and} \quad g_j = \binom{4}{n_j}$$

then $J = 4$ and

$$\bar{x}_j = x_j^0 \equiv g_j^0 (1-x)^{(4-n_j)} x^{n_j} \quad (7)$$

where x_j^0 are the populations of a random alloy. Thus, no matter how large the constant δ may be, as long as the

energies Δ_j vary linearly with the number of B atoms on the cluster n_j and the counting degeneracy is not split, the alloy is random. Strain energies from bond-length differences between the constituents, chemical energies and electron-electron interactions all have some nonlinear variations with n_j . The degeneracy g_j^0 is split by coherent strains, temperature gradients, or anything that establishes preferred directions for the locations of A and B atoms on a cluster. These splittings always establish a preference for one type of cluster in a particular orientation [e.g. for a strain in the (111) direction, A_3B with the B oriented along the (111) axis], and therefore drive deviations of the $\{\bar{x}_j\}$ set from $\{x_j^0\}$ toward compound formation, and, if the splittings are large enough, toward long-range order. The effect of the energies $\{\Delta_j\}$ can drive the $\{\bar{x}_j\}$ set toward either compound formation or spinodal decomposition, depending on details.

The excess free energy of an $A_{1-x}B_xC$ alloy can be written as

$$\Delta F = \Delta E - T\Delta S \quad (8)$$

where the excess energy is

$$\Delta E = M \sum_j \varepsilon_j \bar{x}_j \quad (9)$$

with M being the number of clusters. The excess entropy is

$$\Delta S = k \ln \left(\frac{N!}{N_A! N_B!} \prod_{j=0}^J \frac{M_j!}{M_j!} [\theta_j (1-x)^{4-n_j} x^{n_j}]^{M_j} \right) \quad (10)$$

with $M_j = x_j M$ (see Sher *et al* [21] for the justification). The total number of clusters M is related to the number of Bravais lattice sites N in the crystal by $M = N/4$ for a 16-bond cluster (and $M = N$ for a four-bond cluster). Equations (9) and (10) can be rewritten for the 16-bond cluster as

$$\Delta E = \frac{N}{4} \left((1-x)\varepsilon_0 + x\varepsilon_j + \sum_j \Delta_j \bar{x}_j \right) \quad (11)$$

and

$$\begin{aligned} \Delta S &= -kN \left((1-x) \ln(1-x) + x \ln x \right. \\ &\quad \left. + \frac{M}{N} \sum_j (\bar{x}_j \ln \bar{x}_j - \bar{x}_j \ln x_j^0) \right) \\ &= \frac{N}{4T} \sum_j \Delta_j \bar{x}_j - N \frac{\mu_B}{T} x + \frac{N}{4} \ln q(\{\Delta_j\}, \mu_B). \end{aligned} \quad (12)$$

Note in the first equality in equation (12) that M/N is 1 for a four-bond and 1/4 for a 16-bond cluster. In the 16-bond cluster case, there is, as a consequence, an exact cancellation of the $(1-x) \ln(1-x) + x \ln x$ term by a portion of the third term in the bracket. This occurs physically because, in the 16-bond cluster, all the bonds associated with a given substituted atom are in the cluster. In the four-bond case, however, bonds from each substituted atom contribute to four different clusters. Several features of equations (11) and (12) are noteworthy. First, ε_0 and ε_j are functions of x but are

temperature independent. For bond-length-mismatched alloys, they vary with x roughly as $\varepsilon_0 \approx \varepsilon_0^0 x^2$ and $\varepsilon_j \approx \varepsilon_j^0 (1-x)^2$, so that we have

$$\Delta E \approx N \left(\Omega x(1-x) + \sum_j \Delta_j \bar{x}_j \right) \quad (13)$$

where the mixing enthalpy parameter, Ω is approximately

$$\Omega \approx [\varepsilon_j^0 + (\varepsilon_0^0 - \varepsilon_j^0)x]/4 \quad (14)$$

and nearly x independent if $\varepsilon_0^0 \approx \varepsilon_j^0$, as is often the case. The second term in equation (13) is usually small compared with the first and contains all the temperature dependence. However, when equations (5), (12) and (13) are inserted into equation (8) to obtain ΔF , a term in $T\Delta S$ exactly cancels the second term in equation (13). Thus, it is impossible to determine the temperature variation of ΔE from a measurement of ΔF . The expression for ΔF becomes

$$\Delta F = N[\Omega x(1-x) + x\mu_B - \frac{1}{4}kT \ln q(\{\Delta_j\}, \mu_B)]. \quad (15)$$

The chemical potential μ_B is determined from equation (3), and both μ_B and q are generally temperature dependent.

We have calculated $\{\varepsilon_j\}$, $\{\Delta_j\}$, μ_B , $\{\bar{x}_j - x_j^0\}$, $\Delta E(T)$ and $\Delta F(T)$ for $Hg_{1-x}Cd_xTe$ (figure 2) and $Hg_{1-x}Zn_xTe$ (figure 3) alloys. The energies ε_j for these figures were calculated as discussed in detail in [21]. The 16-bond cluster was attached to a rigid medium at the virtual crystal positions in the third shell from the cluster centre. The atom positions inside the third shell were adjusted to minimize the strain and chemical energies. The energy ε_j is the total energy of the 16 bonds with the atoms in their minimum-energy positions. The variations of atom bond lengths in different alloys are well predicted by this method. The ΔF curves for $Hg_{1-x}Cd_xTe$ show no tendency for an order-disorder transition in the temperature range studied, but $\bar{x}_j - x_j^0$ values do deviate from zero. $Hg_{1-x}Zn_xTe$ does have an order-disorder spinodal transition with critical [36] temperature $T_c \approx 380$ K. Although $Hg_{1-x}Cd_xTe$ shows significant deviations of the \bar{x}_j from random alloy values for a sample equilibrated at 300 K, they are still not as large as those measured in the NMR experiments, meaning that a major interaction may be missing. However, at 300 K diffusion may be fast enough in HgCdTe for the material to equilibrate even at this low temperature (see the discussion in the next section). The slower diffusion observed in HgZnTe may be a distinct advantage, one that counters its above-room-temperature critical temperature.

We have shown that only the small $\{\Delta_j\}$ drive deviations in the populations \bar{x}_j away from randomness. Moreover, if $\Delta_j = n_j \delta$ is linear in n_j , the number of B atoms, then despite the size of δ , the populations are still random. Thus, effects that would otherwise be considered small may compete with the larger energies retained if they have the proper nonlinearity. We are examining several possibilities in $Hg_{1-x}Cd_xTe$. These include effects caused by screening in the composition range where the alloy is a semi-metal, direct second-neighbour chemical

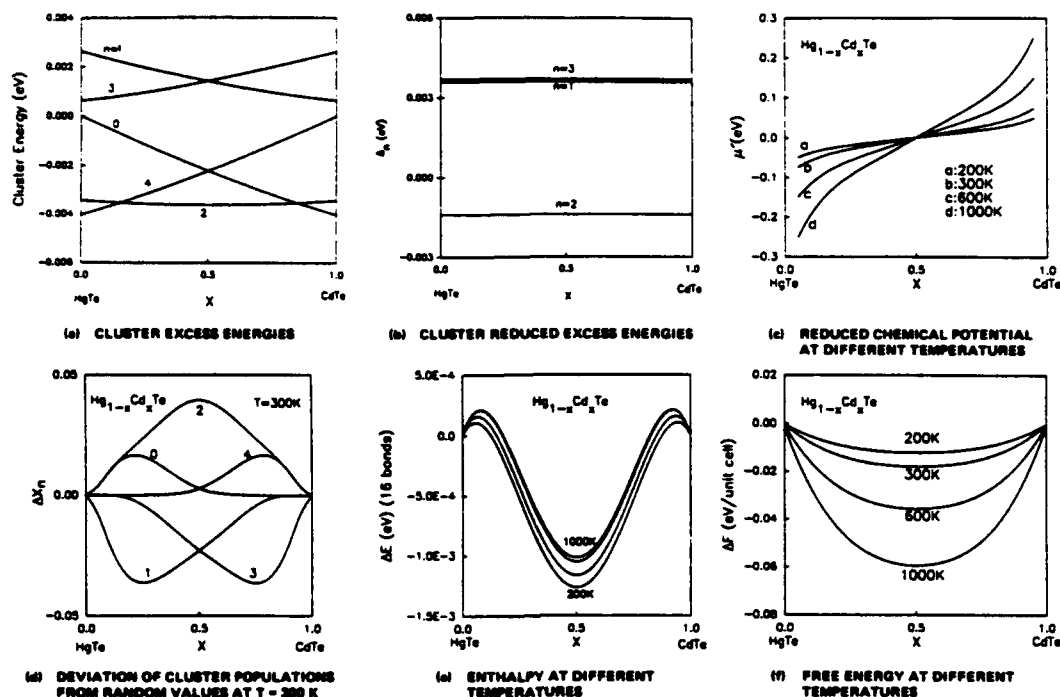


Figure 2. Composition variation of correlation-state-dependent quantities for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$.

interactions treated in the context of Harrison's model [37], electron-electron Coulomb interactions driven by polarity differences between the constituents, and composition variations of the elastic constants.

3. Native defects and diffusion

3.1. Data fitting

There have been several measurements [15, 16] of the Hg vacancy formation energy and Hg, Cd and Te tracer diffusion coefficients. These measurements are always fitted to a functional form

$$G = G_{\infty}(P)\exp(-E'/kT) \quad (16)$$

where E' is interpreted as the activation energy and $G_{\infty}(P)$ is the infinite temperature limit of G . For vacancy formation, G represents the vacancy concentration, and for diffusion the infinite temperature diffusion coefficient. It is assumed in the fitting process that $G_{\infty}(P)$ can be viewed as a function only of the partial pressure P above the sample of the species under investigation. However, the pre-exponential coefficient always has a power-law temperature contribution, i.e. a T^m multiplicative component. Because the measurements extend over a limited temperature range, $T_1 \leq T \leq T_2$, usually have some scatter, and the T^m term varies more slowly than the exponential, reasonable fits to the data are obtained. However, when E' values are quoted to more than one significant figure, and G_{∞} to the proper order of magnitude, it is important to include the T^m factor in the fitting proce-

dure. Because we do not have access to the original data we have corrected published numbers by assuming the straight-line fits of $\ln G$ versus $1/T$ were made through the points T_3 and T_4 given by

$$\begin{aligned} T_3^{-1} &= \frac{1}{3}(2T_1^{-1} + T_2^{-1}) \\ T_4^{-1} &= \frac{1}{3}(T_1^{-1} + 2T_2^{-1}) \end{aligned} \quad (17)$$

Then the correct expression

$$G = C(P)T^m \exp(-E/kT) \quad (18)$$

parameters are related to the fitted parameters by the relations

$$E = E' - mk\tau \quad (19)$$

and

$$\frac{1}{m} \ln \frac{G_{\infty}}{CT_4^m} = \frac{\tau}{T_4} \quad (20)$$

where

$$\tau \equiv (T_3^{-1} + T_4^{-1})^{-1} \ln(T_4/T_3) \cong T_4. \quad (21)$$

The approximate expression to τ in equation (21) is accurate if $T_4 - T_3 \ll T_4$. Note that if m is positive $E < E'$ and if m is negative $E > E'$.

To illustrate the effect of these corrections, we have examined the case of V_{Hg} formation and diffusion and the interstitial Hg_i diffusion in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$. The mass action equation for formation of V_{Hg}



where Hg represents a Hg atom on a Hg site, and Hg_g represents Hg in the vapour phase, has been analysed

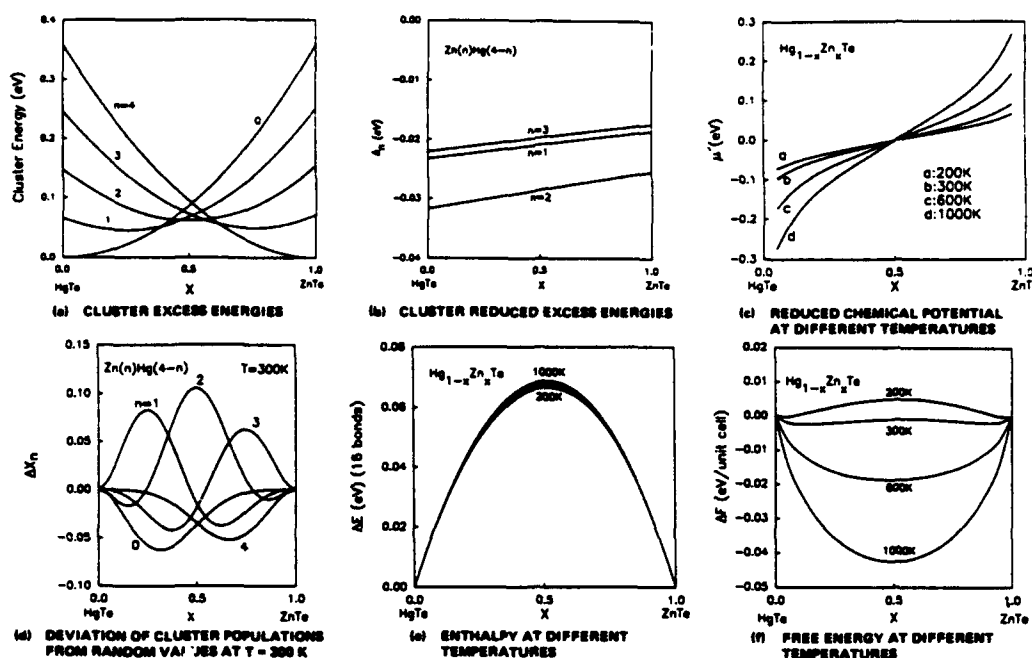


Figure 3. Composition variation of correlation-state-dependent quantities for $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$.

using standard results in statistical mechanics. The diffusion coefficient D is given by

$$D = \frac{1}{6} \frac{a^2}{\tau_c} \left(\frac{n}{N} \right) \quad (23)$$

where

$$\frac{1}{\tau_c} = \omega_p \frac{1}{2} f_I \quad (24)$$

and where a is the hop distance, τ_c is the mean free time between hops, n/N is the fraction of lattice sites N occupied by vacancies, ω_p is the presentation frequency of an atom at the barrier between it and an adjacent vacant site and f_I is the probability that an adjacent atom reaches the intermediate position at the peak of the barrier between it and an adjacent vacancy. (The prime is inserted to remind us that this is not a normal interstitial because it lies between two vacant sites.)

The quantity f_I is calculated roughly from the mass action equation for the effective reaction



again remembering that the HgTe is adjacent to a V_{Hg} in the initial state.

Then, the assumption that the Hg_I in the intermediate position does not remain long enough for its vibrational states to thermally equilibrate leads to the power-law m values quoted in table 1.

There are several noteworthy features to the information in table 1. First, the corrections to the activation energies are 10 to 20%, which is important if the second and third significant figures quoted are to be taken seriously. Second, note that for both vacancy and interstitial diffusion in the Chen [16] and in the Tang and Stevenson [15] work a larger activation energy correlates to a larger pre-exponential factor. This occurs

Table 1. Corrections to experimental activation energies and pre-exponential coefficients.

Experiment	E (eV)	G_∞ ($\text{cm}^2 \text{s}^{-1}$)	T_1 (K)	T_2 (K)	τ	E (eV)	C	m
Vacancy formation								
Vydyanath [13, 14]	2.24		400	650	495	2.01		11/2
Vacancy diffusion								
Chen [16]	2.40	$4.7 \times 10^5 P_{\text{Hg}}^{-1}$	400	490	441	2.08	$3.2 \times 10^{-21} P_{\text{Hg}}^{-1} \dagger$	17/2
Tang and Stevenson [15]	2.10	$4.87 \times 10^4 P_{\text{Hg}}^{-1}$	350	500	412	1.80	$5.8 \times 10^{-22} P_{\text{Hg}}^{-1} \dagger$	17/2
Interstitial diffusion								
Chen [16]	0.54	$1.1 \times 10^{-7} P_{\text{Hg}}$	400	490	441	0.64	$5.5 \times 10^{-13} P_{\text{Hg}} \ddagger$	-5/2
Tang and Stevenson [15]	0.61	$5.5 \times 10^{-7} P_{\text{Hg}}$	350	500	412	0.70	$2.4 \times 10^{-13} P_{\text{Hg}} \ddagger$	-5/2

$\dagger (\text{cm}^2 \text{s}^{-1}) - T^{17/2}$

$\ddagger \text{cm} - T^{5/2} \text{s}^{-1}$.

because of the T^m dependence and the fact that the two experiments were conducted over different temperature intervals T_1 and T_2 .

The corrected numbers for the two groups are closer together than the uncorrected ones in both cases, though still not in perfect agreement. If we average the corrected Hg diffusion activation energies of the two experimental groups we get

$$E_{D,V_{Hg}} = 1.94 \pm 0.2 \text{ eV}$$

$$E_{D,Hg_i} = 0.67 \pm 0.04 \text{ eV}.$$

3.2. Native defects

Much of the following discussion has appeared previously in [2]. As mentioned earlier the doubly ionized cation vacancy is believed to be the dominant native defect in HgCdTe that dictates the electrical behaviour of the undoped material [13, 14]. We have recently reported the results of calculations of native and impurity defect total energies in HgTe, CdTe and ZnTe [18]. The energies for the formation of various native point defects in HgTe are summarized in table 2; the HgTe solid and the Hg in the vapour are used as reference energies†. These calculations used the linearized muffin-tin orbital (LMTO) method within the local density approximation (LDA) to the exchange correlation energy. Large supercells containing one defect per cell were repeated periodically, and from the difference in total energies per cell, with and without the defect, the defect formation energy was calculated. To expedite the calculations, the energies were calculated within the atomic spheres approximation (ASA) with a small (spd) basis set. In the ASA, an approximation to the exact density functional is evaluated; as a result, an error is introduced which is larger than in other LDA methods [38], and relaxation energies cannot be accurately calculated. Thus only those differences in energies that are $> 0.5 \text{ eV}$ here should be viewed as significant for these calculations.

An appropriate set of mass action constants for the neutral defect reactions is also given in table 2. The

† A different reference is used here than was used in table III in [18]. An error appears in that table owing to the incorrect use of an energy of $2E_g$ per unit cell rather than $4E_g$. The defect energies in table I in [18], from which the energies in table III were derived, are correct.

notation in that table is as follows. A_B corresponds to an A species occupying a B site, where I corresponds to an interstitial and V to a vacancy. No subscript on a species indicates that it is occupying the correct lattice site, e.g. $Hg = Hg_{Hg}$. Square brackets [] refer to concentrations. A subscript 'g' indicates the species in the gaseous, or vapour, phase, and P_{Hg} is the Hg vapour pressure. Most of the reactions in table 2 involve the creation or destruction of one or more unit cells. Because the resulting change in volume is accommodated at the surface, the change in the number of unit cells will enter into the determination of the defect equilibrium through the surface entropy. Additionally, surface preparation and orientation will affect the surface free energy. We have assumed for the present that such surface effects are negligible, i.e. that the volume expansions and contractions can occur with negligible changes in the surface properties.

To complete the defect equilibrium determination correctly we must include the equilibration of the electronic charges of the system. To do so we must have knowledge of the dominant charge states of the defects and their activation energies with respect to the neutral defect. Such calculations are complicated by the fact that most *ab initio* calculations of the electronic band structure of semiconductors predict an incorrect bandgap, E_g , a shortcoming of the local density approximation (LDA). Therefore, we shall focus on the neutral native defects here, and the established or expected charge states of these defects.

In wide-bandgap materials the equilibration of defects can be substantially affected by the Fermi level; for example the formation energy of a donor will decrease when the Fermi energy is near the valence band edge, since the donor electron can drop into a vacant state near the valence band, thereby lowering the energy by $\sim E_g$. Because we are discussing HgCdTe with a narrow bandgap, we expect the Fermi effects to be small, but not insignificant at high temperatures. Because of the small conduction band effective mass, in n-type material the filling of the conduction band states by electrons can shift the Fermi energy significantly. Combined with the increase in the bandgap for the high temperatures at which most defect studies are done, the effective bandgap can be substantially larger than the usual 77 K bandgap associated with a given concentration of HgCdTe.

Table 2. Defect reactions and formation energies.

Defect reaction	Defection concentration	Energy (eV)
$E_{V_{Hg}} + HgTe \leftrightarrow V_{Hg}Te + Hg_g$	$[V_{Hg}] = P_{Hg}^{-1} K_{V_{Hg}}^0 \exp(-E_{V_{Hg}}/kT)$	2.01†
$E_{Te_{Hg}} + 2HgTe \leftrightarrow Te_{Hg}Te + 2Hg_g$	$[Te_{Hg}] = P_{Hg}^{-2} K_{Te_{Hg}}^0 \exp(-E_{Te_{Hg}}/kT)$	4.53
$E_{Te_i} + HgTe \leftrightarrow Te_iTe + Hg_g$	$[Te_i] = P_{Hg}^{-1} K_{Te_i}^0 \exp(-E_{Te_i}/kT)$	4.96
$E_{V_{Te}} + Hg_g \leftrightarrow HgV_{Te}$	$[V_{Te}] = P_{Hg} K_{V_{Te}}^0 \exp(-E_{V_{Te}}/kT)$	3.12‡
$E_{Hg_{Te}} + 2Hg_g \leftrightarrow HgHg_{Te}$	$[Hg_{Te}] = P_{Hg}^2 K_{Hg_{Te}}^0 \exp(-E_{Hg_{Te}}/kT)$	-0.42
$E_{Hg_i} + Hg_g \leftrightarrow Hg_i$	$[Hg_i] = P_{Hg} K_{Hg_i}^0 \exp(-E_{Hg_i}/kT)$	0.84, 0.98

† Corrected experimental number from Vydyanath [13, 14]

‡ Calculated using a tight-binding Hamiltonian [17].

First we consider the defects which accommodate excess tellurium—the first three defects in table 2. The mass action constants are given by a product of the form

$$K_i = K_i^0 \exp(-\Delta E_i/kT). \quad (26)$$

For the first three equations the K_i^0 are given by

$$K_{V_{Hg}}^0 = C_0(kT)^{5/2}(2\pi m_{Hg})^{3/2}h^{-3} \exp(\Delta S_{V_{Hg}}/k) \quad (27)$$

$$K_{Te_{Hg}}^0 = C_0(kT)^5(2\pi m_{Hg})^3h^{-6} \exp(\Delta S_{Te_{Hg}}/k) \quad (28)$$

and

$$K_{Te_i}^0 = C_0(kT)^{5/2}(2\pi m_{Hg})^{3/2}h^{-3} \exp(\Delta S_{Te_i}/k) \quad (29)$$

Here T is the temperature in kelvin, k is Boltzmann's constant, m_{Hg} is the mass of the mercury vapour atoms, h is Planck's constant, ΔS_i is the change in vibrational entropy upon formation of the defect and C_0 converts from site fraction to volume concentrations. Estimates, valid at high temperature, of the temperature variation of the ΔS_i terms were included in the pre-exponential dependence T^m of the reaction constants in table 1. Because two unit cells of HgTe are destroyed when a tellurium antisite is created, compared with one unit cell when a mercury vacancy is created, we do not expect that $\exp[(\Delta S_{Te_{Hg}} - \Delta S_{V_{Hg}})/kT] \approx 1$. While we have not completed the evaluation of these entropy terms, our preliminary estimates indicate that this ratio is $\sim 10^4$. For the tellurium interstitial and the mercury vacancy, we expect that $\exp[(\Delta S_{Te_i} - \Delta S_{V_{Hg}})/kT] \approx 1$ will be correct within a factor of 10. Evaluating the numerical constants we find

$$\frac{[Te_{Hg}]}{[V_{Hg}]} \approx 10^{-10} \quad (30)$$

$$\frac{[Te_i]}{[V_{Hg}]} \approx 10^{-18} \quad (31)$$

for $T = 500^\circ\text{C}$ and $P_{Hg} = 1$ atm. The conclusion from equations (30) and (31) that the mercury vacancy is the dominant native defect is consistent with experimental observation. This conclusion is unchanged if we include the possibility that the species may be ionized at the equilibration temperature where the material is expected to be intrinsic. Although the tellurium antisite density decreases more rapidly with decreasing Hg pressure than does the mercury vacancy density, the point at which the concentrations are comparable is at less than $P_{Hg} \approx 10^{-10}$ atm, and certainly the HgTe phase boundary is reached before such low Hg pressures can be achieved. This is also consistent with the fact that no p-to-n conversion is observed in isothermal anneals for low mercury pressures [13, 14, 38], as would be expected if tellurium antisites became the majority native defect. Because the pressure dependences of the tellurium interstitial and the mercury vacancy concentrations are the same, the above conclusions will hold independent of the mercury pressure.

We have checked the sensitivity of the calculated concentration ratios to the magnitude of the reaction enthalpy. Because the enthalpies enter the exponents, small changes in the enthalpies will result in large

changes in the predicted defect concentrations. For example, let us assume that our calculated antisite formation enthalpy is in error by 0.5 eV: in this case the ratio of antisite-to-vacancy concentrations (at 500°C) will be reduced to $\sim 1 \times 10^{-11}$. For an antisite formation enthalpy in error by 1.0 eV, this ratio is reduced to $\sim 1 \times 10^{-7}$. We do not expect the ASA errors to exceed 0.5 eV [5].

If the HgCdTe is not completely annealed, and tellurium precipitates are still present, the defect equilibrium will not be that predicted by the mass action equations given in table 2. For example, near the inclusions we can assume that the defects will be nearly in equilibrium with the tellurium solid; thus



and



will be the appropriate reactions. The formation energies for a tellurium antisite and an Hg vacancy from the tellurium solid are calculated to be 1.63 eV and 1.15 eV, respectively. Although the difference in the formation energies is less than when both defects are referenced to the mercury vapour (~ 0.5 eV compared with ~ 2 eV), the gas phase entropy factor does not enter into the ratio of the defect concentrations. Using the same estimate of the entropy ratio, the defect concentration ratio using tellurium solid as the reference state is

$$\frac{[Te_{Hg}]}{[V_{Hg}]} \approx 10^{-8} \quad (34)$$

Thus, near the inclusions we expect higher relative concentration of tellurium antisites, as compared with the rest of the material equilibrated with the Hg vapour. Additionally, the absolute $[V_{Hg}]$ defect concentrations may differ substantially in the two regions of the material. At present we expect $[V_{Hg}]$ to be higher in an 'atmosphere' surrounding a Te inclusion. A better calculation of the vibrational entropy is needed before we can predict these absolute defect concentrations and their spatial extent. Differences in the defect concentrations arising from different equilibration conditions are a possible source of spatial variability of the HgCdTe material. If the material is not fully annealed to equilibrium, for example because of an abundance of tellurium precipitates, this history may affect subsequent processing.

In the above we have discussed the defect energies for HgTe and applied them directly to the small-x $Hg_{1-x}Cd_xTe$ system. Because we are dealing with the native defects of an alloy, we expect a number of complexities to affect the above analysis. First the variation of the defect formation energies for vacancies is sensitive to the alloy environment, in particular for the vacancies of the non-substituted species, such as tellurium in HgCdTe [39]. Even for vacancies of the substituted species, we have found that the formation energies may vary by several tenths of an electron volt. Because of this variation in the formation energy, the fraction of defective sites

will vary by as much as a factor of 100 from one class of sites to the next. Consider various classes of Hg sites in ideal HgCdTe, which can be distinguished to first order by specifying the number of Hg and Cd atoms in the second-neighbour shell (the four first neighbours are always tellurium), $\text{Hg}_{12-j}\text{Cd}_j$, with a concentration given by $[j]$. The total vacancy concentration is given by

$$[V_{\text{Hg}}] = \sum_{j=1}^{12} [j] P_{\text{Hg}}^{-1} K_j^0 \exp(-E_j/kT) \quad (35)$$

where E_j is the vacancy formation energy for the j th cluster. The populations of vacancies in each class of cluster, j , can be expected to differ because of differences in the cluster populations and the formation energies. Additionally, the defect energy levels may differ in the various classes of sites, possibly leading to different ionization states for vacancies in different classes of sites. If the cations in the alloy are randomly arranged, such differences may be difficult to infer experimentally. If, however, the cations are correlated, exhibiting short-range order, more complex behaviour may be present. As discussed earlier, such short-range order has been found in HgCdTe. In these cases, the contribution to the vacancy densities from the dominant class of clusters will be increased. Because the studies finding short-range order focus on the tellurium-centred five-atom clusters of the form $\text{Hg}_4\text{-Cd}_n$ rather than on cation-centred clusters of the form $\text{Hg}_{12-n}\text{Cd}_n$, higher-level five-atom cluster-cluster correlations must be known to predict the effects on the vacancy populations.

Next we examine the defects which accommodate excess Hg in the solid. The existence region for HgCdTe is always tellurium rich, and thus the native defects which accommodate excess tellurium are expected to dominate. For these equations in table 2, K_i^0 is given by

$$K_{V_{\text{Te}}}^0 = C_0^{-1} (kT)^{-5/2} (2\pi m_{\text{Hg}})^{-3/2} h^3 \exp(\Delta S_{V_{\text{Te}}}/k) \quad (36)$$

$$K_{\text{HgTe}}^0 = C_0^{-1} (kT)^{-5} (2\pi m_{\text{Hg}})^{-3} h^6 \exp(\Delta S_{\text{HgTe}}/k) \quad (37)$$

and

$$K_{\text{Hg}_i}^0 = C_0 (kT)^{-5/2} (2\pi m_{\text{Hg}})^{-3/2} h^3 \exp(\Delta S_{\text{Hg}_i}/k). \quad (38)$$

If we assume the change in entropy is comparable for all three defects, we find

$$\frac{[\text{Hg}_{\text{Te}}]}{[V_{\text{Te}}]} \approx 10^{+14} \quad (39)$$

and

$$\frac{[\text{Hg}_i]}{[V_{\text{Te}}]} \approx 10^{-14} \quad (40)$$

for $T = 500^\circ\text{C}$ and $P_{\text{Hg}} = 1$ atm. From equations (39) and (40) we see that the tellurium vacancy is a minority defect species. For the pressure and temperature considered, the density of Hg antisites is predicted to be comparable to the density of Hg interstitials. Because the ratio of $[\text{Hg}_{\text{Te}}]$ to $[\text{Hg}_i]$ is nearly unity, any errors in the calculation of the activation energy could push the balance toward one side or the other. Thus we must

depend on the next generation of calculations, with the ASA removed and full relaxation included, plus a quantitative comparison of the entropy differences between the mercury antisite and the tellurium interstitial, to determine the dominant defect in this class.

As mentioned above, we have shown that the tellurium vacancy formation energy varies significantly with the alloy environment. Because the tellurium vacancy is not expected to be a dominant defect in HgCdTe, and the tellurium diffuses by an interstitial mechanism, we do not expect any measurable manifestation of this variation. On the other hand, the Hg antisite may be the dominant Hg-excess defect, and its formation energy may vary significantly with the alloy environment. We are currently calculating the magnitude of this variation.

3.3. Diffusion

HgCdTe exhibits a complex tracer diffusion profile, with both a fast and a slow branch. The fast branch is attributed to a mechanism with vacancy and interstitial diffusion in parallel where the dominant diffuser is determined by the pressure and temperature, while the slow component fits a mechanism with vacancy and interstitial diffusion in series [15]. The activation energies for the fast branch as discussed previously are 1.90 eV and 0.67 eV for the vacancy and interstitial mechanisms, respectively. Our calculated formation energy for the mercury interstitials are 0.84 and 0.98 eV for the anion- and cation-surrounded tetrahedral interstitial sites, respectively, and the experimental formation energy for the mercury vacancy is 2.01 eV. Comparing these energies with the experimental activation energies we find close agreement, indicating that the migration energy contribution to the diffusion activation energies are small for both interstitials and vacancies.

In the vacancy diffusion case we are comparing two experimental numbers for the diffusion and formation energies. Because the diffusion energy (1.94 eV) is the sum of the formation and migration energy, the fact that the vacancy formation energy (2.01 eV) is larger but should be smaller is an indication that something is amiss, and the experiments should be repeated with the goal of attaining higher accuracy.

There is a similar discrepancy in the interstitial diffusion case, but now we are comparing the experimental diffusion activation energy 0.67 eV with a theoretical formation energy 0.89 eV. We know that refinements to the theory will lower the predicted value. These refinements need to be done before more definitive conclusions can be made.

The difference between the two numbers obtained for cation- and anion-surrounded interstitial sites (0.84 eV and 0.98 eV) sets a lower bound on the interstitial migration energy. Most of the interstitial Hg will sit on the lower-energy anion-surrounded site, and migrate through the intermediate-energy cation-surrounded sites. Examination of the lattice arrangement between these sites leads us to believe that the potential profile is

unlikely to have a large barrier between the two classes of site, and, as a consequence, the interstitial migration barrier should be quite low.

In a recent experiment on mercury diffusion in ion-implantation-damaged HgCdTe, an activation energy of several tenths of an electron volt was measured [40]. The disparate results can be interpreted as a measure of only the defect migration contribution to the diffusion activation energy, since defects in excess of the equilibrium concentration were probably formed during implantation. It is not evident that the measured activation energy corresponds to the vacancy or the interstitial mechanism. The conclusion that the diffusion activation energies are largely defect formation energies, with the migration energies being much smaller, is in agreement with the above interpretation of the Richter and Kalish [40] experiment.

4. Conclusions

We have incorporated our calculated defect energies into the mass action equations for the neutral defects in HgCdTe. In agreement with experiment, we find the mercury vacancy to be the dominant native defect in tellurium-rich material. We also find the mercury antisite and interstitial defect densities to be comparable, although a better calculation of the vibrational entropy is needed to confirm this result. Comparing the defect formation energies of Vydyanath [13, 14] and our theory to the diffusion measurements by Chen [16] and by Tang and Stevenson [15], we find agreement with their diffusion activation energies for both the vacancy and the interstitial mechanisms, if we assume that the migration energy is small in both cases. Small migration energies are consistent with the measurement of Richter and Kausli [40] and our theory. Further work is in progress to incorporate the defect charge states into the calculation, and also to calculate the fully relaxed defect energies with the full potential LMTO.

If the cation migration energies are as small as we are suggesting, then some cation motion will be present even at room temperature. This implies that the crystal can equilibrate to its room temperature order-disorder correlation state. Most semiconductor alloys at room temperature are in metastable correlation states frozen to the temperature where diffusion stopped as they were cooled. How this affects devices remains to be determined.

We predict that near a Te inclusion there will be an 'atmosphere' of defects differing from the one the bulk solid has in equilibrium with a Hg v. air. In this atmosphere the ratio of the concentrations of the tellurium antisites to Hg vacancies will be increased. The tellurium antisite concentration may still be too small to affect any device properties, but if the Hg vacancy concentration is greater near the inclusions than in the bulk, this will be a source of spatial variability in the material. This is a potential source of the worms in LPE material. We are cautious about this suggestion because

TEM studies of annealed LPE material indicate that there are few, if any, Te inclusions remaining [7].

First-principles theories are now fast enough and accurate enough for practical use in process and performance modelling. An effort to capture the results of many experiments into consistent models, and thereby increase their reliability, has, as we have tried to demonstrate in this paper, become a realistic goal.

Acknowledgment

This work was supported by NASA contract NAS1-18226, by ONR contract N00014-88-C0096 and by AFOSR contract F49620-88-K-0009.

References

- [1] Balcerak R 1991 *Semicond. Sci. Technol.* **6** C1
- [2] Berding M A, Sher A and Chen A-B 1991 *Mater. Res. Soc. Symp. Proc.* **216** 3
- [3] Sher A, Berding M A, Chen A-B and Patrick R 1988 *Advanced Research Workshop NATO International Science Exchange Programme, 5-9 Sept 1988*
- [4] Sher A, van Schilfgaarde M and Berding M A 1991 *J. Vac. Sci. Technol.* **B 9** 1738
- [5] Chen M C, Parker S G and Weirauch D F 1985 *J. Appl. Phys.* **58** 3150
- [6] Yoshikawa M, Maruyama K, Saito T, Maekawa T and Takigawa H 1987 *J. Vac. Sci. Technol.* **A5** 3052
- [7] Schaake H F, Tregilgas J H, Lewis A J and Everett P M 1983 *J. Vac. Sci. Technol.* **A1** 1625; Schaake H F, Tregilgas J H, Beck J D, Kinth M A and Gnade B E 1985 *J. Vac. Sci. Technol.* **A3** 143
- [8] Nguyen C, Triboulet R and Lemasson P 1987 *Seventh American Conf. on Crystal Growth in Conjunction with II-VI-87, 12-17 July, Monterey CA*
- [9] Patrick R S, Chen A-B, Sher A and Berding M A 1989 *Phys. Rev. B* **39** 5980
- [10] Spicer W E, Silberman J A, Lindau I, Chen A-B, Sher A and Wilson J A 1983 *J. Vac. Sci. Technol.* **1** 1735
- [11] Cole S, Carey G P, Silberman, J A, Spicer W E and Wilson J A 1985 *J. Vac. Sci. Technol.* **A 3** 206
- [12] Cooper D E and Harrison W A 1990 *J. Vac. Sci. Technol.* **A 8** 1112
- [13] Vydyanath H R 1981 *J. Electrochem. Soc.* **128**, 2609
- [14] Vydyanath H R, Donovan J C and Nelson D A 1981 *J. Electrochem. Soc.* **128** 2625
- [15] Tang M-F S and Stevenson D A 1990 *J. Vac. Sci. Technol.* **7** 544 and references therein
- [16] Chen J S 1985 *PhD Thesis* University of Southern California
- [17] Jones C E, James K, Merz, J, Barunstein R, Burd M, Eetemadi M, Hutton S and Drumheller J 1985 *J. Vac. Sci. Technol.* **A3** 131
- [18] Berding M A, van Schilfgaarde M, Paxton A T and Sher A 1990 *J. Vac. Sci. Technol.* **A 8** 1103
- [19] Destefanis G L 1991 *Semicond. Sci. Technol.* **6** C88
- [20] Sher A, Chen A-B and van Schilfgaarde M 1986 *J. Vac. Sci. Technol.* **A 4** 1965
- [21] Sher A, van Schilfgaarde M, Chen A-B and Chen W 1987 *Phys. Rev. B* **36** 4279
- [22] Zamir D, Beshah K, Becla P, Wolff P A, Griffin R G, Zax D, Vega S and Yellin N 1988 *J. Vac. Sci. Technol.* **6** 2612

- [23] Amirtharaj P M and Pollak F H 1984 *Appl. Phys. Lett.* **45** 789
- [24] Compaan A, Bowman R C and Cooper D E 1990 *Appl. Phys. Lett.* **56** 1055
- [25] Vodopyanov L K, Kozyrev S P, Aleshchenko Y A, Triboulet R and Y Marfaing 1990 *Appl. Phys. Lett.* **56** 1057
- [26] Quintana J P 1990 private communication (Mater. Res. Soc. Boston)
- [27] Shin S, Arias J, Zandian M, Pasko J and DeWames R private communication
- [28] Johnson S M, Rhiger D R, Rosbeck J P, Peterson J M, Taylor S M and Boyd M E, private communication
- [29] Chandra D private communication;
Chandra D, Tregilgas J H and Goodwin M W 1991 *J. Vac. Sci. Technol. B* **9**, 1852
- [30] Muller M W 1984 *Phys. Rev. B* **30** 6196
- [31] Parayanthal P and Pollak F H 1984 *Phys. Rev. Lett.* **52** 1822
- [32] Sher A, Berding M A and Chen A-B 1989 *J. Cryst. Growth* **98** 27
- [33] Wei S-H and Zunger A 1988 *J. Vac. Sci. Technol. A* **6** 2597
- [34] Czyzyk M T, Podgorny M, Balzarotti A, Letardi P, Motta N, Kisiel A and Zemmal-Slarnavska M Z 1986 *Z. Phys. B* **62** 3850
- [35] Ichimina M and Sasaki A 1986 *J. Appl. Phys.* **60** 3850
- [36] Chen A-B and Sher A 1985 *Phys. Rev. B* **32** 3695
- [37] Harrison W A 1980 *Electronic Structure and the Properties of Solids* (San Francisco: Freeman)
- [38] Jones C L, Quelch M J T, Capper P and Gosney J J 1982 *J. Appl. Phys.* **53** 9080
- [39] Berding M A, Sher A and Chen A-B 1990 *J. Appl. Phys.* **68** 5064
- [40] Richter V and Kalish R 1990 *J. Appl. Phys.* **67** 6578

Appendix H

JONES' THEORY OF THE HUME-ROTHERY PHASES REVISITED

Jones' Theory of the Hume-Rothery Phases Revisited

A.T.Paxton¹ M. Methfessel² and D.G.Pettifor³

¹SRI International, Menlo Park, California 94025

²FHI, Faradayweg 4-6, D-1000 Berlin 33, Germany

³Department of Mathematics, Imperial College, London SW7 2BZ

1. INTRODUCTION

It is usual to divide metal alloy phases into three classes: electrochemical compounds, size factor compounds and electron compounds.¹ (The term "compound" is used rather broadly here to include secondary solid solutions.) Electrochemical compounds are those that form following classical rules of chemical bonding in which the coordination numbers reflect the valencies of the pure components. Size factor compounds embody the famous Hume-Rothery "15% rule", and include substances such as Laves phases and other topologically close-packed structures whose atoms are of largely differing sizes and have therefore only limited ranges of solid solubility. An intriguing principle in the theory of alloys is the abundance of alloys in the third category, the Hume-Rothery "electron compounds": there appears a strong correlation between the number of valence electrons per unit cell, valence electron concentration or electron per atom ratio and the crystal structure adopted by certain alloy phases, be it fcc, bcc, hcp, β -Mn, γ - or ϵ -brass structures. For example, the appearance of the β -phase at the solid solubility limit of primary α -alloys of noble metals with elements in groups 12, 13 and 14 mostly occurs at a valence electron concentration around 1.4. In the light of modern density functional theories of electronic structure, it seems all the more remarkable that such a simple rule can hold true over ranges of alloys that comprise simple, noble and transition metals. However, the phenomenological evidence is hard to dispute: both the large number of electron compounds, and the behaviour of defect structures. For example, non-stoichiometric γ -brass will incorporate vacancies seemingly in order to maintain a valence electron concentration that corresponds to the electron per atom ratio of 21/13 which stabilises the complex cubic structure. NiAl has an ordered bcc structure with an electron/atom ratio of 3/2 (transition metals are assigned a valency of zero in the Hume-Rothery theory); in compounds containing less than 50% Ni, some of the Ni-sites remain vacant and it has been suggested^{1,2} that this is to maintain a valence electron concentration favorable to the formation of bcc phases. It is possible that this phenomenon has nothing to do with the valence electron concentration. Instead it may reflect the large formation energy of an Al antisite defect in the highly ordering Ni-Al system compared with the vacancy formation energy. In fact we believe that the Hume-Rothery e/a rule is unlikely to be operative, for example, in Ni-Al or Ni-Zn alloys, and that since the rule was stated many merely coincidental observations have been attributed to it. In the case of noble metal alloys with groups 12, 13 and 14, however, we shall show that the e/a rule can be seen as a direct consequence of the general validity of the rigid band model, as first suspected by Jones³ in 1937.

It has been the aim of many theoretical studies over the past 55 years to explain the electron compounds in terms of electron theory; good and comprehensive reviews of this effort can be found in recent works by Cottrell⁴ and Hafner.⁵ A very detailed discussion of the pseudopotential approach can be found in Heine and Weaire.⁶ A detailed review of experimental data and rigid-band models has been made by Massalski and Mizutani.⁷ Many studies have concentrated on the archetypal system Cu-Zn, since this reveals most of the principles of the theory, and also since this was the subject of the pioneering study of the electron compounds³ resulting in the famous "Jones Theory of Alloying." A glance at the phase diagram (see figure 7a) shows that while Zn is hcp and has a wide range of solubility in Cu, which is fcc; between the solubility limits $0.3 < x < 0.6$ $\text{Cu}_x\text{Zn}_{1-x}$ alloys contain bcc β -brass. Thus as one adds Zn to Cu, varying the valence electron concentration, n , continuously between one and two, the structure of the alloys changes in the sequence $\text{fcc} \rightarrow \text{bcc} \rightarrow \text{hcp}$, the phase transitions occurring at approximately $n = 1.4$ and $n = 1.8$ (we omit discussion of the complex cubic γ -brass phase occurring between $n = 1.6$ and $n = 1.8$). The purpose of Jones' theory was to explain the transition from fcc to bcc at a valence electron concentration of 1.4.

Jones' theory was based on the following three assumptions.

1. The energy difference between phases is given entirely by the difference in the band-structure energy in the two crystal structures. In other words, given a bandstructure for the crystalline solid (that is to say, the relationship $\epsilon(k)$ between wavenumber k and electron eigenenergy) and its associated density of states $g(E)$, the bandstructure energy is

$$\begin{aligned} E_{\text{band}} &= \sum_{i \text{ occ.}} \epsilon(k_i) \\ &= \int_{-\infty}^{\epsilon_F} \epsilon g(\epsilon) d\epsilon, \end{aligned} \quad (1)$$

where the sum is taken over occupied electronic states. The Fermi ϵ_F energy (maximum occupied energy level) is determined by the valence electron concentration n according to

$$n = \int_{-\infty}^{\epsilon_F} g(\epsilon) d\epsilon. \quad (2)$$

2. In order to compute the energy difference, one can assume energy bands which deviate from free-electron behaviour only as they approach Bragg reflecting planes in the Brillouin zone and only in the direction normal to these planes. Only under these conditions could the electron dispersion $\epsilon(k)$ be written in closed form.
3. The rigid band approximation. This states that the effect of alloying is to add or subtract electrons without altering the bandstructure. In other words, we can assume that the bandstructures of all $\text{Cu}_x\text{Zn}_{1-x}$ alloys are same: they only differ in the positions of their Fermi energies. A consequence of the rigid band approximation is the implication that alloys are disordered at all compositions. This will also be implicit in our work, so that we are not able to account for the ordering of the β -phase into the β' (CsCl) structure. Both these two approximations can be removed within the Coherent Potential Approximation and General Perturbation Method.⁸ As far as the density of states is concerned, this is not necessary in the case of Cu-Zn, where the rigid band approximation has shown to be valid in neutron scattering experiments.⁹ On the other hand, in the Cu-Ni system our approach would not be

justified.⁹ We will also find that the rigid band model severely overestimates the heat of formation of β -CuZn. In spite of these reservations we will keep to the rigid band approximation so that we are able to interpret our results within the framework of Jones' theory of alloying. Also, since the phase diagram shows no α/β equilibrium at 0°K, our purpose here is only to expose general principles concerning the Hume-Rothery e/a rule.

Recent advances are able to address the first two assumptions more clearly than previously. The assumption 1 is now known to be rigorously provable to first order in density functional theory, and is known as the structural energy difference theorem.¹⁰ As Jones pointed out,¹¹ if assumption 2 is not invoked, then the evaluation of the density of states function becomes a formidable task. (Indeed even then, a further approximation needed to be made by Jones:³ we will return to this in section 4.2.1 .) This is particularly so, since the Jones theory is centred in the nature of the density at and near Van Hove singularities and these are difficult to reproduce even with modern methods of Brillouin zone integration. Our present paper is devoted to an examination of the theory of Cu-Zn alloys in which we apply the structural energy difference theorem to self-consistent rigid band models. Using an advanced technique for Brillouin zone integration, we explicitly show the Van Hove singularities responsible for the transitions from fcc to bcc to hcp as the valence electron concentration is varied. The structure of the paper is as follows. In section 2 we describe Jones' seminal work, its popularisations, its relation to other current theories and its interpretation in terms of general arguments involving typical shapes of the density of states. In section 3 we show how two recent advances, namely the structural energy difference theorem and the analytic quadratic method of integration allow us to remove all of Jones' assumptions except the rigid band approximation for the alloy. In section 4, we make density functional calculations of phase stability in the Cu-Zn system using both the self-consistent bandstructure in the local density approximation, and various model bandstructures in an attempt to recover Jones' original theory. We confirm the validity of the rigid band approximation in this context, and we are able to make detailed statements about the way in which the positions of Van Hove singularities in the bandstructure drive the transitions between phases. In section 5 can be found a summary and our conclusions.

2. THEORIES OF THE ELECTRON COMPOUNDS

2.1 The Mott and Jones, and Jones theories

Unfortunately, considerable confusion has accumulated in the metallurgical literature surrounding Jones' theory of β -phase alloys. This situation arises from the fact that two quite different explanations of the Hume-Rothery rule are attributed to Jones. We will call these the Mott and Jones, and the Jones models and will describe them separately in the following two paragraphs. The situation has been very clearly unravelled by Hume-Rothery.¹² We also refer the reader to discussions in Barrett and Massalski,¹³ and Heine and Weaire.⁶

Initially, Hume-Rothery¹⁴ observed that in a large number of binary compounds of univalent fcc metals, the phase diagram displayed a limit of solubility ($\alpha/(\alpha + \beta)$ phase boundary) at a valence electron concentration of about 1.4 . Furthermore the appearance of a single bcc β -phase ($(\alpha + \beta)/\beta$ phase boundary) occurred at a valence electron

concentration around 1.5 . Mott and Jones¹⁵ advanced the following explanation for the phenomenon. If we imagine a free-electron Fermi sphere expanding inside the fcc Brillouin zone as the valence electron concentration is increased between one and two, we find that the sphere makes contact with the {111} zone faces at a valence electron concentration of 1.36 . Similarly in the bcc Brillouin zone the sphere will first contact the {011} faces when the valence electron concentration reaches 1.48 . Mott and Jones pointed out that the density of states would go through a maximum at the point where the Fermi sphere made contact with the zone boundaries. This is because Bragg-reflection causes otherwise free-electron-like bands to deviate and become flat. This will lead to a cusp-like peak in the density of states. If, furthermore, we compare two structures, one of whose density of states is a free-electron parabola, while the other has a maximum due to this zone boundary effect, then one can argue qualitatively that when the Fermi energy lies on the maximum, the latter structure will be stable since electrons can be accommodated under the peak in the density of states at lower energies than in the competing structure. Therefore as the valence electron concentration in an fcc solid solution is increased from one, at a value of 1.36 the density of states in the stable fcc phase will begin to decrease and that phase will become less stable than a competing structure with free-electron-like density of states since the decreasing density of states will force electrons into higher energy states and de-stabilise the fcc phase. This is said to explain the formation of β -phase alloys at valence electron concentrations around 1.4 . At a valence electron concentration of 1.48, the bcc phase will be most stable; this is the explanation for the ubiquitous occurrence of β -phase alloys at the famous Hume-Rothery electron/atom ratio of 3/2. Clearly, the Mott and Jones model implies that at a valence electron concentration of one (that is to say, in Cu) the Fermi surface is not in contact with any of the faces of the fcc Brillouin zone. Measurement of the Fermi surface of Cu by Pippard in 1957 showed that this was incorrect, and that the Fermi surface of Cu formed elongated necks at the {111} Brillouin zone boundaries.¹⁶ This is often quoted as indicative of the complete failure of Jones' theory. An important fallacy in the Mott and Jones theory is that if the bands were free-electron-like so that contact with the zone boundaries was made at the free-electron values of 1.36 and 1.48, then there would be no peaks in the density of states.¹² Conversely, if there were peaks (arising from deviations from free-electron behaviour) then they would not occur at the free-electron values but at lower values of valence electron concentration. A final consequence of the popularisation of the Mott and Jones model is the statement encountered in several elementary texts that stability of a structure arises when the Fermi level falls in a *maximum* in the density of states, and that a decreasing density of states indicates instability with respect to a competing phase. This rather counter-intuitive statement is indeed incorrect, as we shall show in section 2.2 .

The above model is frequently referred to as the Jones theory of alloying, even in modern metallurgical textbooks.¹⁷ But Jones' work of 1937 is distinguished from the Mott and Jones model in two respects. Firstly, Jones explicitly incorporated the large band gap observed on the Bragg-reflecting {111} planes in fcc Cu; and secondly, rather than making qualitative arguments Jones numerically integrated his model density of states to distinguish the energies of the fcc and bcc phases. The characteristic peaked structure of the fcc and bcc densities of states is shown in figure 1a. These were obtained by Jones from approximate Brillouin zone integrals of a model dispersion for Cu which included a band gap of 4.1eV deduced from optical measurements. The effect of the large gap is to move the peaks at the zone boundaries to valence electron concentrations much lower than the

free-electron values. In fact the Fermi level in Cu indicates that in Jones' model the Fermi surface was about to make contact with the zone boundaries (see figure 1a). Numerical integrations of the densities of states puts the $\alpha|(\alpha + \beta)$ and $(\alpha + \beta)|\beta$ phase boundaries at valence electron concentrations of 1.41 and 1.45 in very good agreement with what was known of the phase diagram in 1937. These values do not, of course, correspond with the peaks in Jones' densities of states, which occur at approximately 1 and 1.2 electrons/atom.¹² In fact fcc is at its most stable at a valence electron concentration corresponding to the peak in the bcc density of states (see figure 1b). This result is completely at odds with the conventional wisdom since at this point the fcc density of states is rapidly decreasing; however we shall see in section 2.2 that this is in fact quite consistent with the shapes of the densities of states. Much of the confusion in the literature, in fact, has arisen from a mixing up of the Mott and Jones, and Jones models. In most textbooks^{1,17} we find Jones' figure (our figure 1a) reproduced, but with the peaks in the density of states labelled with the free-electron values 1.36 and 1.48 electrons per atom. Furthermore, it is claimed that fcc becomes unstable with respect to bcc as soon as the fcc density of states begins to decrease. This is not Jones' result as can be seen from figure 1b; however, it is not immediately obvious since the density of states is plotted against energy whereas the structural energy differences are plotted against valence electron concentration. It is often clearer to show both quantities plotted against n . Then for a given bandfilling, one can immediately compare the densities of states and the energy difference.

Heine and Weaire⁶ have pointed out the important distinction between theories that attribute special stability to structures in which contact is made between a Brillouin zone boundary and the Fermi surface and those that involve contact with the Fermi sphere, which is the Fermi surface the alloy would have if the bands were truly free-electron like. There is no justification in the former case, as can already be seen from Jones' results and will be demonstrated below. The Mott and Jones assertion that stability arises when the Fermi sphere makes contact is phenomenologically correct, and we will see why it approximately works in section 2.2. This is also the result that emerges from second order perturbation theory because the leading terms in these theories are indeed response functions of a free-electron gas which have weak singularities at twice the free-electron Fermi wavevector. However, these singularities do not immediately lead to minima in the bandstructure energy. Therefore even in pseudopotential perturbation theory there is no complete explanation of the Hume-Rothery rule of electron compounds. It will be necessary to accurately calculate energy bands and carefully sum the occupied states. This is the approach we shall adopt here. There are two advantages over pseudopotential perturbation theory, on which attempts to understand the Hume-Rothery rule has invariably been based in the past.^{5,6} Firstly, our bandstructure energies are exact to all orders; and secondly we will properly include sd -hybridisation in our calculations. We will show explicitly that neglect of d -band effects leads to an incorrect prediction of the structure of pure Cu.

2.2 General arguments

Following Jones¹⁸ and Massalski and Mizutani,⁷ it is possible to make some quite general arguments in the context of which we will be able to make an analysis of our results below. At the same time the shortcomings of the Mott and Jones theory become immediately evident.

Given the density of states of two different structures as function of the energy, it is difficult to envisage the difference of the bandstructure energies as function of the valence electron concentration, since the interplay between two different Fermi levels must be considered. In order to understand the behaviour of the bandstructure energy, one may exploit the fact that the derivatives of E_{band} with respect to the valence electron concentration from equations (1) and (2) are given by

$$\frac{dE_{band}}{dn} = \epsilon_F \quad (3)$$

$$\frac{d^2E_{band}}{dn^2} = \frac{1}{g(\epsilon_F)}. \quad (4)$$

Taking the difference of eq. 3 for two different structures, we see that the derivative of the energy difference is zero when the two Fermi energies are equal. At these values of n , the difference of the two bandstructure energies ΔE_{band} is therefore extremal (except in the special case of a saddlepoint). To determine whether the extremum is a maximum or a minimum, we take the difference of eq. (4) between the two structures. It follows that the more stable structure at the extremum is the one with the lower density of states at the Fermi energy.[†] This suggests the following procedure for analysing the energy difference between two phases. One should first search for those n for which the Fermi energies are equal and the energy difference is extremal. This is easily done by varying a common Fermi energy until the occupied parts of the two densities of states have equal areas. By comparing the two densities of states at the Fermi energy, the more stable structure can be determined. The critical valence electron concentrations for the crossover between phases (where the two bandstructure energies are equal) can be estimated from the positions of the extrema.

We are now in a position to point out the basic fallacy in the Mott and Jones theory. The central argument of the theory is that the instability of a phase is associated with the steep drop behind a peak in the density of states (which is correct) and that the instability occurs immediately after states behind the peak start filling up (which is incorrect). To see the error, consider figure 1c which shows the density of states for three typical cases. The first, labelled "P", is the free-electron parabola. The second, labeled "F" shows the typical behaviour when the free-electron bands are perturbed by a set of Bragg planes. The third curve, "B", is similar to curve F, but corresponds to a Bragg plane further from the origin. For the Jones model of the Cu-Zn $\alpha \rightarrow \beta$ transition, F and B would correspond to fcc and bcc, respectively. Comparing phases F and P, it is seen that the two Fermi energies are equal at energy ϵ_1 , at which curve F rejoins the free-electron parabola, as well as all higher energies. Note that the triangular areas above and below the free-electron parabola must be equal. The two densities of states are equal at ϵ_1 , so that the above criterion for deciding between a maximum or minimum does not apply.

[†] The converse applies, however if ΔE_{band} is a minimum and greater than zero or a maximum and less than zero. Under these circumstances the most stable structure will be that having the highest $g(\epsilon_F)$. This case will not arise in the present context, but in other situations such as application to the energy ordering in transition metal trialuminides¹⁹ we have found exactly this state of affairs. In that case a naive appraisal of the densities of states would have led to an incorrect structure prediction.

However, curve F has clearly moved electrons to lower energies and is more stable. In fact, phase F has a lower bandstructure energy at all n , in disagreement with the statement that F should become unstable as soon as the peak at point A is passed.

Turning to the more relevant case of phases F and B , one sees that equal density of states areas and the extremal bandstructure energy difference occur at the energy labelled ϵ_2 . There, the density of states of phase F lies below that of B , so that phase F is more stable. We see that phase F does in fact have a lower energy when the Fermi energy reaches point A, as discussed by Mott and Jones,¹⁵ but then becomes ever more stable with increasing valence electron concentration until point ϵ_2 is reached. The reason is easy to understand. We start from the situation in which F should be the most stable according to the Mott and Jones model, namely when the F and B Fermi energies lie at point A and at some higher energy, respectively. Incrementing the electron number in small steps, each extra amount of charge enters at a lower Fermi energy for F than for B and serves to further stabilize phase F . This continues well into the range where the F -density of states lies below the B -density of states, until finally the two Fermi energies are equal and the maximal bandstructure energy difference is reached. Consequently, and contrary to the Mott and Jones model, the crossover at which phase B becomes stable relative to F must lie at an even higher valence electron concentration.

These examples have demonstrated two interesting facts. Firstly, the phase transition does not happen as soon as states after the peak at point A are filled, but comes considerably later. Secondly, unless the density of states of the second phase deviates from the free-electron parabola, there will be no transition at all. Therefore the critical valence electron concentration always depends on the shape of both densities of states, and cannot be deduced from prominent features of the first alone. We can now come to conclusions concerning the question of the Fermi surface or Fermi sphere touching the Brillouin zone face. Clearly there is no special significance to the valence electron concentration at which the Fermi surface touches; this is just point A, and all we know is that the transition is somewhere higher up. However, it can be expected that the valence electron concentration for which the free-electron sphere touches will be a useful guide in many cases. This is simply because two effects tend to cancel: the peak at point A always lies below this free-electron valence electron concentration, while the energy crossover will lie a comparable amount above the peak. Given a sequence of phases as in the Cu-Zn system, the critical valence electron concentrations deduced from the free-electron sphere will probably give the correct sequence of the phases and a reasonable estimate of the phase boundaries.

It is obvious from the common tangent construction that the position of the phase boundaries must depend on the free energies of both phases as emphasised in the last paragraph. This point seems to have been neglected in the literature. In the simplest (Mott and Jones) model, one considers the critical valence electron concentration at which the fcc phase becomes less stable than bcc; in other words, the concentration at which the energies of the two phases are equal. This has no special significance in determining phase boundaries; except that the $\alpha|(\alpha + \beta)$ and $(\alpha + \beta)|\beta$ boundaries must bracket this point.

3. THE DENSITY FUNCTIONAL APPROACH

3.1 Frozen potential calculations

Although Jones³ had to take it as an assumption that the structural energy difference was just the difference in bandstructure energies, we now know that in density functional theory there is a rigorous proof of this to first order.¹⁰ Suppose we want the difference in total energy between two monatomic crystals with the same atomic volume but having different crystal structures, I and II. Then we may make a calculation of the self-consistent potential of structure I and the bandstructure energy eq. (1) associated with that potential. Then, if we have a way of rigidly transferring this potential into structure II, we may calculate the bandstructure energy of that structure in the *frozen potential*. Then the total energy difference between structures I and II is given by the difference in their bandstructure energies. This can be shown to be an exact result to first order in the difference between the frozen potential in structure II and its self-consistent potential.^{10,20} Of all the bandstructure methods available to density functional theory, the LMTO-ASA method²¹ most lends itself to this kind of calculation since the potential can be expressed completely in terms of a set of four "potential parameters" for each inequivalent atom and for each orbital type: *s*, *p* and *d*. The frozen potential approach can be more accurate than a self-consistent LMTO-ASA calculation. This is because the major source of error in LMTO-ASA arises from the spheridisation of the charge density in calculating electrostatic terms in the total energy. In the structural energy difference theorem these terms cancel to first order and so do not contribute any error.

Our procedure for calculating structural energy differences is the same as that described by Skriver.¹⁰ (See also van Schilfgaarde *et al.*²²) In the present paper, this method shall be called the local density approximation-rigid band (LDA-RB) approach. We have chosen to do all our calculations at the measured atomic volume of β -CuZn (see table 1). We begin by making a self-consistent LMTO calculation of fcc Cu, including "combined correction" and augmenting up to $\ell_{\max} = 2$, i.e., including *s*, *p* and *d* orbitals in the basis. Combined correction²¹ is a means of going beyond the atomic spheres approximation (ASA) in the LMTO method. By adding extra terms in the hamiltonian and overlap matrices the combined correction accounts for those regions of space in which the atomic spheres overlap, or which are not included in any atomic sphere, by computing integrals between unaugmented LMTO functions in these regions. The combined correction also includes unaugmented basis functions from $\ell_{\max} + 1$ to ∞ inside the atomic spheres. Although the combined correction somewhat detracts from the simple elegance of the ASA and structural energy difference theorem, it is found to be necessary in computing the very subtle energy differences between fcc and bcc noble metals. Essentially our approach is to take a trial potential within which we calculate energy bands as accurately as possible. Our self-consistent potential parameters are shown in table 1. This potential is then used to generate bandstructures for fcc, bcc and hcp Cu. The only differences between the structures then appear in the structure constant and combined correction matrices. (This is another elegant feature of the "spherical" methods such as LMTO in which the scattering properties of the atomic potentials separate out from the structure of the lattice in the secular equations.)

From the bandstructures, we generate densities of states, which we then integrate up to Fermi energies corresponding to a number of electrons per atom which we vary

in increments between 11 (Cu) and 12 (Zn). In this way, we are using a rigid band formulation to the alloy problem to investigate the relative stability of fcc, bcc and hcp as a function of number of electrons. We plot this as two energy differences, bcc-fcc and hcp-fcc against valence electron concentration.

3.2 Reconstructing the density of states

Because the whole problem, and particularly the interpretation, hinges on the existence of Van Hove singularities in the bandstructure at the Brillouin zone boundaries, we have devoted some effort to the problem of obtaining the density of states from the bandstructure $\epsilon(k)$ where we evaluate the eigenvalues ϵ of the hamiltonian at a discrete set of reciprocal lattice vectors $\{k\}$.

This is a perennial problem in solid state theory: namely, how to construct a density of states function from eigenvalues given on a mesh of k -points. Solutions fall generally into two classes: those that use only the information given at each of the discrete set (these are called "sampling" methods); and those which interpolate the bandstructure within the k -mesh and then calculate the density of states analytically from the interpolated bandstructure. Of the latter class, the most widely used is the linear tetrahedron method of Jepsen and Andersen.²³ Here, the reciprocal unit cell is divided into tetrahedra with apexes lying on the k -mesh. The bands are then interpolated linearly within each tetrahedron resulting in a continuous (but not differentiable) bandstructure everywhere within the unit cell. For this approximate bandstructure, the density of states is calculated analytically. A serious drawback is that the interpolation will lead to errors if there are bands that cross within a tetrahedron. Also neither linear nor sampling methods can resolve Van Hove singularities properly. These are kinks, or points of infinite derivative in the density of states which are caused by k -points at which the gradient of the energy band disappears. Therefore, since the linear terms vanishes at these points, one must interpolate the bands at least to quadratic order to obtain a density of states function with analytically correct Van Hove singularities. This can be done using the "analytic quadratic method"²⁴ in which the k -mesh is used to construct a network of tetrahedra which have a k -point at the mid point of each edge as well as at each apex. This allows the eigenvalues to be interpolated to quadratic order so that the resulting bandstructure is both continuous and once differentiable. Unfortunately this approach still leads to errors from band crossings inside the zone. This problem is largest for the complex of the fine d -bands, which in our case are completely filled. On the other hand it is well known that the best way to obtain integrated quantities over a full band is to use sampling over a regular mesh.²⁵ Therefore in view of the very small energy differences we are interested in, we have adopted the following strategy. To obtain the most accurate integrals over the d -bands, we have used a sampling method over a mesh constructed by marking 40 divisions along the three reciprocal lattice basis vectors. For displaying densities of states, and for integrals over purely s - and p - bands, we have used the analytic quadratic method using eigenvalues from 11700, 9139 and 19635 k -points in the irreducible wedges of the fcc, bcc and hcp Brillouin zones respectively. For easy reference in what follows, we show in figure 2 the fcc, bcc and hcp Brillouin zones and the standard labelling of the symmetry points and lines.

4. RESULTS AND DISCUSSION

In this section, we show our solution to the Cu-Zn problem. Firstly, we calculate structural total-energy differences using the structural energy difference theorem described in the previous section. Having demonstrated that the LDA-RB approach gives the correct answers, we then attempt by successive approximations to recover the Jones, and Mott and Jones theories within our framework.

4.1 The local density approximation-rigid band results

The self-consistent LMTO bandstructures of fcc Cu and Zn are shown in figure 3 alongside the densities of states calculated using the analytic quadratic method. These are energy bands calculated within the local density approximation to density functional theory. We have used the parameterisation of the von-Barth-Hedin exchange and correlation energy density used by Moruzzi *et al.*²⁶ The self-consistent potential parameters are given in table 1. To generate the bandstructures of bcc and hcp Cu the potential parameters from self-consistent fcc Cu are used and only the structure constants and combined correction differ, as described in section 3.1. The resulting bandstructures are shown in figure 4, where we concentrate on the bands around the Fermi energy lying above the *d*-bands. Figure 5 shows the density of states of fcc and bcc Cu superimposed. In this figure the Van Hove singularities are labelled with symbols referring to the Bragg-reflecting Brillouin zone boundary planes which have produced the singularities and the symmetries of the corresponding wavefunctions. As indicated in figure 2, L and X refer to the fcc {111} and {002} planes, while N refers to the {110} planes in the bcc zone. They are clearly seen in figure 4 as energies at which the slope of the energy bands vanishes.

Let us briefly discuss the important features in figures 3, 4 and 5. We first note that the difference between Cu and Zn lies primarily in the position of the *d*-bands which in Zn are well below the Fermi energy, while in Cu the *sp*-bands around ϵ_F are clearly perturbed by the *d*-bands which are rather shallow. We note that the Fermi energy in Cu lies between the bands labelled L_2 and L_1 . L_2 is a pure *p* state and the gap between it and the pure *s* L_1 state above it is caused by the requirement that the free-electron-like *s* and *p* states be orthogonal to the tightly bound *d*-states.²⁷ The position of the L_1 state is further raised due to "repulsion" from the lowest *d* state which also has L_1 symmetry.²⁸ This is the origin of the measured gap of 4.1eV used in Jones' calculations.³ We note that the effect of *sd*-hybridisation is to produce an asymmetric band gap at the Brillouin zone boundary. Normally in perturbation theory the effect of Bragg-reflection is to split the energy evenly about the free-electron value. (Asymmetric splitting is a common feature in the bandstructures of noble metals.) As expected the gap becomes smaller in the absence of strong hybridisation with the *d*-band as is seen to be the case in Zn (figure 3). In the Mott and Jones picture, there are no *d*-bands, the gap at L is smaller and the Fermi energy lies below the L_2 *p*-state. We see that the real life case is rather far removed from the Mott and Jones model. The origin of the raised Fermi energy is the shrinking of the constant energy surface as *sd*-hybridisation is turned on, the *sp*-bands bending over to join into the highest lying *d*-bands. This can be seen clearly in the Γ to K panel in figure 3. The effect of *sd*-hybridisation is to bend the uppermost *sp*-band in towards Γ so that at a given energy, the surface of constant energy shrinks. Therefore in order to accommodate a given number of electrons the Fermi energy must be raised.

We now consider the structural energy differences arising from application of the

structural energy difference theorem and the rigid band approximation. In all that follows we will use results based on the self-consistent fcc Cu potential. However as a rather extreme test of the rigid band approximation, we show in figure 6 the structural energy differences ΔE_{band} computed using both the self-consistent Cu and self-consistent Zn potentials given in table 1. Using the Cu potential certainly bears out the observed phase transitions in the Cu-Zn phase diagram. In order to assess the validity of the structural energy difference theorem in the atomic spheres approximation, we also show in figure 6 the energy differences in pure Cu and Zn calculated using a full-potential LMTO method.²⁹ We believe that these results are essentially exact within the local density approximation. It is notable that using the Cu potential one can reproduce very closely the phase ordering in Zn. This shows that in this case the rigid band approximation is not very severe. It is important to note that the Zn potential leads to the hcp structure being slightly more stable in Cu. This is because the Zn potential produces relatively low-lying *d*-bands, so that the *sd*-hybridisation effects, that we will see below are so important in stabilising Cu in the fcc structure, are not reproduced.

It is clear that the shift of the Fermi energy to a position above the L_2 singularity, and the distortion of the free-electron-like nature of the bands due to the shallow *d*-bands means that a different interpretation of the phase transitions in the Cu-Zn system than the traditional one is called for. In order to consider the LDA-RB results in the light of the general considerations of section 2.2, we show in figure 8 bandstructure energy differences from figure 6 and the density of states and difference in Fermi energy $\Delta\epsilon_F$ plotted as functions of *n*. (With inclusion of the *d*-bands, note that the total number of electrons per atom is $n + 10$.)

To begin with, we can see the origin of the stability of fcc Cu in the local density approximation. Since the Fermi energies at $n = 1$ are almost equal (-0.133 , -0.134 and -0.135 Ry in fcc, bcc and hcp respectively), we can conclude that the stable phase of Cu will be that having the lowest $g(\epsilon_F)$. This is fcc as seen in figures 5 and 8. In fact we can attribute the stability of fcc with respect to bcc to the shapes of the respective Brillouin zones. We see from figure 5 that the fcc density of states is low between the L_2 and X_4 singularities, while the bcc density of states is raised near the Fermi energy by the singularity at N_1 .

At higher valence electron concentrations, the fcc density of states is raised above the bcc by the X_4 singularity, and we see in figure 8c that shortly after this singularity is reached the bcc and fcc Fermi energies become equal leading to a minimum in the bcc-fcc energy difference and the stabilisation of the β -phase.

In the light of our remarks in section 2.2, we can see that the competition between fcc and bcc structures can be analysed by identifying those valence electron concentrations at which the Fermi energies are equal. This occurs at two energies, indicated in figure 8c and by vertical lines in figure 5. At these points, fcc and bcc respectively are at their most stable, the stable phase being that with the lowest density of states at the Fermi energy.

We also conclude that hcp is stabilised above $n = 1.6$ due to the L_1 and N_1 singularities at the top of the band gap in the fcc and bcc structures. This effect is discussed in more detail in section 4.2.1. The β -phases are stabilised by the steady linear decrease in $g(\epsilon)$ in the bcc band gap, while the fcc density of states encounters a further singularity at the {200} zone faces. Rather than the L_2 singularity at the {111} faces of the Brillouin zone being responsible for destabilisation of the fcc phase as in the Mott and Jones theory, in

the LDA-RB model this peak is washed out by the shoulder from the d -bands, and the additional faces of the Brillouin zone at $\{200\}$ are responsible for the $\alpha \rightarrow \beta$ transition. Furthermore, we have seen that features at the various Brillouin zone faces have conspired with the steeply falling shoulders from the d -bands to favour the fcc phase in Cu. We will see below that this result can only be obtained using realistic bandstructures (*i.e.*, explicitly including the d -bands).

From the data of figure 6a, we have computed free-energy-composition curves in order to estimate the zero-temperature phase boundaries. These curves are shown in figure 7b, sheared according to Zener's device³⁰ in order to allow the common tangents to be drawn. Our $\alpha|(\alpha + \beta)$ and $(\alpha + \beta)|\beta$ phase boundaries are very closely spaced at around 1.33 electrons per atom. This construction is only useful in a comparison with Jones' results^{3,30} which we defer until section 4.2.1. The zero temperature phase boundaries are unknown. Indeed, it is reported³¹ (see figure 7a) that β' -CuZn decomposes eutectoidally at 250°C, so that the true limit of solubility can only be found if the equilibrium with the γ phase is taken into account.[†] This has also been neglected by Turchi *et al.*⁸ who however find a zero-temperature solubility limit at $n = 1.25$ which differs from our $n = 1.32$. The discrepancy must arise from our use of the rigid band approximation, which is also inappropriate for the determination of the heat of formation of CuZn. In our LDA-RB model we obtain a heat of formation for disordered β -CuZn of -50kJ/g-atom . This greatly overestimates the experimental heat of formation³¹ of $-9.4 \pm 1.25\text{kJ/g-atom}$; and our calculated value of -12.9kJ/g-atom for ordered β' -CuZn using full-potential LMTO.²⁹ We have independently estimated³² the enthalpy change upon ordering to be of the order of -4kJ/g-atom (which compares well with a value of -3.5kJ/g-atom given by Turchi *et al.*⁸), so that the heat of formation is properly reproduced in the local density approximation, identifying the rigid band approximation as the principal source of error.

4.2 Return to the Simple Theory

We are now at the point where we have applied Jones' methods to the self-consistent bandstructure in the local density approximation in the Cu-Zn system. We have confirmed that indeed the observed phase transitions are closely linked to the bandstructure (as is to be expected in the light of the structural energy difference theorem) and we have confirmed that the LDA-RB approach predicts the correct sequence fcc \rightarrow bcc \rightarrow hcp as the valence electron concentration increases between one and two. But we have found the shallow d -band in Cu to be an important influence in the problem. Firstly we find that the features of the bandstructure thought to be responsible for the fcc \rightarrow bcc transition (namely the positions of the L_2 and N_1 Van Hove singularities) are below the

[†] In which case, the common tangent being drawn between α and γ curves, the $\alpha|(\alpha + \beta)$ phase boundary would move further to the left as seen in the experimental phase diagram. The phase diagram indicates that β' -CuZn (like NiTi) is unstable below a temperature as high as 600°C. We have in fact found³² that in the local density approximation β' -CuZn has a zero value for its elastic shear constant $\frac{1}{2}(c_{11} - c_{12})$. In a similar calculation, NiAl is found to have a negative value for this elastic constant. This is to be expected since the bcc structure is unstable with respect to this shearing in the absence of covalent or ionic bonding.³⁰ For this reason β' -CuZn is only stabilised at elevated temperature by a large vibrational entropy, and is not stable at 0°K.

Fermi energy and largely masked in the upper shoulder of the d -electron density of states. Secondly, we have come to suspect that the stability of fcc Cu is in fact due to the effects of the d -band. In fact Hafner⁵ has pointed out that a general feature of pseudopotential perturbation theories is that there is no ground state fcc phase at a valence electron concentration of one. This he attributes to the neglect of sd hybridisation (see also Heine and Weaire⁶).

Our next step is to make some *model* potentials designed to reproduce the bandstructures of simpler theories. We consider two models in this section. In the first, we wish to explicitly separate out the effect of the d -bands while retaining all the other features of the bandstructure in the local density approximation. This model will represent rather closely the Jones theory. In order to represent the Mott and Jones theory we need a nearly free-electron model in which the effect of the d -bands is ignored and the band gaps are smaller than those due to sd non-orthogonality and band repulsion.

4.2.1 The Jones Model

We seek a model potential with which to apply the structural energy difference theorem and which as closely as possible mimics the Jones model described in section 2. We achieve this using a procedure in which we re-calculate the potential in the atomic sphere in the configuration $4s4p4d$ rather than $4s4p3d$, i.e., constraining the d -wavefunction to have one extra node. The s and p potential parameters are therefore unchanged and d parameters are those of $4d$ rather than $3d$ electrons. This moves the d -electrons out of the conduction band. Figure 9 shows the bandstructures and densities of states of fcc, bcc and hcp Cu in the model potential. This closely reproduces the sp bands of figure 3 while leaving out the d -electrons, so that we can separate out the d -band effects. The structural energy differences are shown and analysed in figure 10. The principal feature of the electron compounds appears, namely the transitions to bcc and then hcp. However, the stability of Cu at $n = 1$ is in doubt; in fact Cu is found to be bcc, with only a small region of fcc stability around $n = 1.3$.

We need to understand why, although ours and Jones' model give similar looking densities of states, the stabilities at $n = 1$ are different. The reason lies in the shapes of the density of states well below the Fermi energy and in the approximation Jones' used in his $\epsilon(k)$ dispersion and in calculating integrals under the Fermi surface.³ The conventional wisdom^{1,17} has it that given two densities of states with two peaks due to Van Hove singularities as in figures 1 and 10, at valence electron concentrations below the peak in the density of states of the structure whose peak comes first, that structure will be stable. This is only true if the two singularities give rise to peaks of similar height. Both ours and Jones' calculations show that the bcc Van Hove singularity is "stronger" than the fcc (i.e., it makes a greater distortion from a free-electron density of states). This means that the bcc density of states is raised above the free-electron density of states by a greater amount than the fcc and this effect persists to energies well below the singularity. The consequence is that bcc is the stable phase over a wide range of valence electron concentration between 0.2 and 1.1. The reason that Jones did not reveal this lies in his approximate dispersion which only deviated from free-electron behaviour close to the Brillouin zone boundaries. It is also possible as suggested by Heine and Weaire⁶ that Jones' approximate method of Brillouin zone integration (the "cone approximation")^{3,7} led to errors.

The relative "strengths" of the first Van Hove singularities in the sequence hcp : fcc : bcc seen in figure 10 arise simply from the numbers of equivalent Bragg planes contributing; these are respectively the $\{11\bar{2}0\}$ prism planes in hcp, the $\{111\}$ planes in fcc and the $\{110\}$ planes in bcc of which the numbers of variants occur in the ratio 6 : 8 : 12.

As in the LDA-RB model, we can apply the analysis of section 2.2 to the results shown in figure 10. We recall that one can predict the structure of the ΔE_{band} curves from the Fermi energy differences. At the points where $\Delta\epsilon_F$ is zero one can predict either a maximum or a minimum in ΔE_{band} depending on whether $\Delta\epsilon_F$ is decreasing or increasing. We can also make the connection between Van Hove singularities and zeros in the Fermi energy difference as follows. At a Van Hove singularity the Fermi energy does not increase with small increases in n . Now, we are treating fcc as the reference phase here, so if the Van Hove singularity is in the bcc or hcp density of states then $\Delta\epsilon_F$ will decrease rapidly and if $\Delta\epsilon_F$ is initially positive this can lead to a maximum in ΔE_{band} (i.e., fcc stability) if the Fermi energy difference falls to zero. If a Van Hove singularity appears in the fcc density of states then $\Delta\epsilon_F$ will rise rapidly and if the Fermi energy is higher in the fcc phase then $\Delta\epsilon_F$ may rise through zero leading to a minimum in ΔE_{band} , i.e., instability of the fcc structure. This effect can be clearly seen in figure 10. At the bcc and hcp Van Hove singularities $\Delta\epsilon_F$ falls rapidly through zero leading in turn to maxima in ΔE_{band} . The fcc singularity at L_1 causes sharp increases in both bcc and hcp $\Delta\epsilon_F$; only in the hcp case does this cause $\Delta\epsilon_F$ to go through zero—this is the origin of the minimum in the hcp-fcc ΔE_{band} curve and the stability of hcp Zn. Again, this is associated with a low hcp density of states when the fcc and bcc density of states rise rapidly through the L_1 and N_1 Van Hove singularities.

We can now see the origin of the stability of bcc and hcp in our reproduction of Jones' model. In the region of valence electron concentration between 1.2 and 1.7 the fcc density of states is steadily decreasing, whereas there appear large peaks in the bcc and then hcp density of states. These correspond to zeros in $\Delta\epsilon_F$ and maxima in ΔE_{band} : first in the bcc-fcc curves and then in hcp-fcc. This accounts for the successive stability of fcc and bcc. The hcp phase is stabilised when the bcc and fcc densities of states rise sharply when the Fermi surface encounters the Brillouin zone boundaries at N and L at the top of the large band gap.

4.2.2 The nearly-free-electron (Mott and Jones) model

Finally, we would like to re-create the very simplest model, namely the nearly-free-electron case, first proposed by Mott and Jones and usually described in textbooks as the rationale behind the theory of the stability of β -phase alloys. We require this to be a model without any effect of d -electrons and having small gaps at the Brillouin zone boundaries. We can generate a bandstructure displaying these features using the potential parameters given in table 1. The resulting bandstructure and densities of states are shown in figure 11. The band gaps are now similar to those in a nearly-free-electron metal such as Al. The structural energy differences and their analysis are shown in figure 12. We conclude that the nearly-free-electron picture is not able to predict a transition to the bcc β -phase. In fact Cu is predicted to be hcp as is found in pseudopotential perturbation calculations, and this is seen to be due to the free-electron sphere first making contact with a Brillouin zone face (at M) in the hcp structure. Note that in the real life bands of figure 4, the effect of the singularity and small band gap at M is to deplete the density of states in

the hcp phase making lcp the least stable phase at low valence electron concentration. In the nearly-free-electron case, because the Fermi surface at $n = 1$ is well inside the first Brillouin zone in fcc and bcc structures, the hcp phase is stable in Cu although the band gap at M is small.

5. SUMMARY AND CONCLUSIONS

We have attempted to clear up some of the confusion surrounding the theory of electron compounds. This has involved making self-consistent bandstructures in the local density approximation and using these in conjunction with the structural energy difference theorem and the rigid band approximation to address the problem from a rigorous point of view. We have then systematically removed, firstly the explicit appearance of the d -bands and secondly the very large band gaps at the Brillouin zone faces in order to recover two of the earlier theories.

Within the rigid-band approximation the structural energy difference theorem tells us that the bandstructure is responsible for energy differences if the bands are generated from frozen potentials. Then it is obvious that within this first-order approach the shapes of various Brillouin zones must eventually be determining phase stability. This fact, coupled with our plausibility arguments at the end of section 2.2 gives in our view a general explanation of the Hume-Rothery electron compounds of noble metals alloying with elements from groups 13, 14 and 15. The appearance of rather constant valence electron concentrations at the phase boundaries in these alloys then arises as a consequence of the approximate validity of the rigid band model and the similar features in the bandstructures of the noble metals (in particular they all have a Fermi energy above the L_2 singularity).

Our other conclusions are as follows.

1. In the specific case of Cu alloys, the sequence of phase transitions as the valence electron concentration varies between one and two can be attributed to features in the electronic bandstructure. These are not the features that have previously been thought responsible, namely the singularities arising when the Fermi surface first makes contact with the Brillouin zone boundaries. Hybridisation with the d -bands moves the Fermi energy well above these singularities even in pure fcc and bcc noble metals. The fcc ground state of Cu is a direct consequence of the presence of the shallow d -band. As shown in section 4.1, both the stability of fcc Cu and the transitions to bcc and then hcp can be attributed to specific features arising from the shapes of the different Brillouin zones. It is notable that the β -phase is stabilised as a result of the additional {200} Brillouin zone planes arising in fcc in contrast to the bcc Brillouin zone which only has one kind of face. Stabilisation of hcp above $n = 1.6$ arises as the Fermi energy moves above the bcc and fcc band gaps.
2. The predictions of the Jones model are qualitatively correct, although the positions of the phase boundaries are too sensitive to the details of the model bandstructure for this to be a quantitative theory. Jones' successful predictions of both the experimental phase boundaries and the fcc ground state of Cu must be regarded as fortuitous accidents. In the former instance, the LDA-RB approach gives much worse agreement with experimental phase boundaries than the much more approximate Jones' theory. The fcc ground state of Cu cannot be obtained in an sp -only model so that Jones'

result must come from approximations either in the model dispersion or in the method of Brillouin zone integration.

3. The nearly-free-electron model is unable to predict the correct sequence of phase transformations. Therefore while the description of the theory of electron compounds given in elementary texts is qualitatively simple and pleasing, it cannot be demonstrated by an explicit calculation.
4. The popular theory attributing stability to a peak in the density of states is incorrect. The effect of a peak in the density of states due to a Van Hove singularity is to produce a rapid change in the difference in Fermi energies between that phase and a competing one. This may cause $\Delta\epsilon_F$ to go through zero in which case the bandstructure energy difference will be extremal: whether a maximum or minimum depends which phase initially has the higher Fermi energy (in other words, whether $\Delta\epsilon_F$ is falling or rising). In the cases where ΔE_{band} is extremal and greater than zero if a maximum and less than zero if a minimum, the stable phase will be the one with the smaller density of states at the Fermi energy; furthermore, the two phases will have equal Fermi energies at the extremum in bandstructure energy difference. Only under these restrictive conditions can the densities of states at the Fermi energy be used as a measure of structural stability.

6. REFERENCES

1. A. H. Cottrell, *Theoretical Structural Metallurgy*, 2nd Ed. (Edward Arnold, London, 1955)
2. W. Hume-Rothery and G. V. Raynor, *The Structure of Metals and Alloys*, 3rd. Ed. (Inst. Metals, London, 1954)
3. H. Jones, *Proc. Phys. Soc.*, A49, 250 (1937)
4. A. H. Cottrell, *Introduction to the Modern Theory of Metals*, (Inst. Metals, London, 1988)
5. J. Hafner, *The Structure of Binary Compounds*, eds. F. R. de Boer and D. G. Pettifor, (North-Holland, Amsterdam, 1989) p. 147
6. V. Heine and D. Weaire, *Solid State Phys.*, 24, 250 (1970)
7. T. B. Massalski and U. Mizutani, *Prog. Mat. Sci.*, 22, 151 (1978)
8. P. E. A. Turchi, M. Sluiter, F. J. Pinski, D. D. Johnson, D. M. Nicholson, G. M. Stocks and J. B. Staunton *Phys. Rev. Lett.*, 67, 1779 (1991)
9. S. Berko, *Noble Metal Alloys*, eds. T. B. Massalski et al., (1986)
10. D. G. Pettifor, *Commun. Phys.*, 1, 141 (1977); H. Skriver, *Phys. Rev. B*, 31, 1909 (1985); N. E. Christensen, *Phys. Rev. B*, 32, 207 (1985)
11. H. Jones, *The Theory of Brillouin Zones and Electronic States in Crystals*, (North-Holland, Amsterdam, 1960)
12. W. Hume-Rothery, *J. Inst. Metals*, 90, 42 (1961)
13. C. Barrett and T. B. Massalski, *Structure of Metals*, 3rd. Ed. (Pergamon, Oxford, 1980) p. 346
14. W. Hume-Rothery, *J. Inst. Metals*, 35, 295 (1926)
15. N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys*, (Clarendon Press, Oxford, 1936)
16. A. B. Pippard, *Phil. Trans. Roy. Soc.*, A250, 325 (1957)

17. R. E. Smallman, *Modern Physical Metallurgy*, 4th Ed. (Butterworth, London, 1985);
R. E. Reed-Hill, *Physical Metallurgy Principles*, 3rd Ed. (PWS-Kent, Boston, 1991)
18. H. Jones, *J. Phys. Radium, Paris*, **23**, 637 (1962)
19. A. T. Paxton and D. G. Pettifor, submitted to *Scripta Metall. Mater.*
20. W. M. C. Foulkes and R. Haydock, *Phys. Rev. B*, **39**, 12520 (1989)
21. O. K. Andersen, *Phys. Rev. B*, **12**, 3060 (1975); O. K. Andersen, NATO ASI on
Electronic Structure of Complex Systems (Ghent 1982), ed. P. Phariseau and W.
M. Temmerman (Plenum, New York, 1984) p. 11; O. K. Andersen, O. Jepsen and
M. Šob, *Electronic Band Structure and its Applications*, ed. M. Yussouff (Springer,
Berlin, 1987)
22. M. van Schilfgaarde, A. T. Paxton, A. Pasturel and M. Methfessel, *Mat. Res. Soc.
Symp. Proc. Vol. 186, Alloy Phase Stability and Design*, ed. G. M. Stocks, A. F.
Giamei and D. P. Pope (MRS, Pittsburgh, 1990)
23. O. Jepsen and O. K. Andersen, *Solid State Commun.*, **9**, 1763 (1971); G. Lehmann
and M. Taut, *phys. stat. sol. (b)*, **54**, 469 (1972)
24. M. Methfessel, M. H. Boon and F. M. Mueller, *J. Phys. C*, **16**, L949 (1983); M. H.
Boon, M. Methfessel and F. M. Mueller, *J. Phys. C*, **19**, 5337 (1986); M. Methfessel,
M. H. Boon and F. M. Mueller, *J. Phys. C*, **20**, 1069 (1987)
25. M. Methfessel and A. T. Paxton, *Phys. Rev. B*, **40**, 3616 (1989)
26. V. L. Moruzzi, J. F. Janak and A. R. Williams, *Calculated Electronic Properties of
Metals*, (Pergamon, New York, 1978)
27. D. G. Pettifor, *Phys. Rev. B*, **2**, 3031 (1970)
28. B. Segall, *Phys. Rev. Lett.*, **7**, 154 (1961)
29. M. Methfessel, *Phys. Rev. B*, **38**, 1537 (1988); M. Methfessel, C. O. Rodriguez and
O. K. Andersen, *Phys. Rev. B*, **40**, 2009 (1989); A. T. Paxton, M. Methfessel and H.
M. Polatoglou, *Phys. Rev. B*, **41**, 8127 (1990)
30. C. Zener, *Phys. Rev.*, **71**, 846 (1947)
31. R. Hultgren, P. D. Desai, D. T. Hawkins, M. Gleiser and K. K. Kelley *Selected Ther-
modynamic Properties of Binary Alloys* (ASM, Ohio, 1973)
32. A. T. Paxton and M. Methfessel, unpublished work

TABLE 1: LMTO potential parameters for self-consistent Zn and Cu and for Cu in the Jones, and Mott and Jones models (see text). C and Δ are the band centre and width parameters; γ measures the distortion from canonical bands and $1/\sqrt{p}$ is the third order parameter whose magnitude indicates the range of energies about the centre of the occupied band in which the linear method is accurate. v_{mtz} is the muffin-tin zero and is needed to construct the combined correction matrix. The Wigner-Seitz radius is 2.748 bohr. γ is a number; all other quantities are in Rydbergs.

			C	Δ	γ	$1/\sqrt{p}$	v_{mtz}
self-consistent	Zn	s	-0.5347	0.1442	0.4129	3.3266	
		p	0.3983	0.1371	0.1078	4.9128	
		d	-0.5868	0.0054	-0.0005	0.4899	-0.7257
self-consistent	Cu	s	-0.4280	0.1527	0.4205	3.7893	
		p	0.5143	0.1465	0.1113	5.4501	
		d	-0.3132	0.0078	-0.0031	0.5614	-0.6937
Jones model	Cu	s	-0.4280	0.1527	0.4205	3.7893	
		p	0.5143	0.1465	0.1113	5.4501	
		d	2.0776	0.1247	0.0437	4.0860	-0.6937
Mott & Jones model	Cu	s	-0.4630	0.1324	0.4000	2.3829	
		p	0.6251	0.1501	0.1125	5.6633	-0.6036

FIGURE CAPTIONS

FIGURE 1: illustrates Jones' theory. (a) The density of states of fcc and bcc Cu. The vertical line shows the Fermi energy in fcc Cu in Jones' model. (b) The fcc-bcc energy difference as a function of valence electron concentration calculated by Jones. Note that the peak in the fcc density of states occurs at $n \approx 1$, and in the bcc at $n \approx 1.2$. (a) and (c) after Jones³ (c) Qualitative sketch of typical densities of states. Energies ϵ_1 and ϵ_2 mark the common Fermi energy for phases F , P and F , B respectively, for which the bandstructure energy difference is extremal.

FIGURE 2: (a) fcc, (b) bcc and (c) hcp Brillouin zones. (After Jones¹¹)

FIGURE 3: Self-consistent local density approximation bandstructures in fcc Cu and Zn. To the left are shown the corresponding densities of states. Note how the analytic quadratic method clearly picks out the Van Hove singularities in the bandstructure. Here and in the next two figures the density from the d -bands is not shown in full as it rises to nearly 100 states/Ry/spin (see Moruzzi *et al.*²⁶), dwarfing the structure relevant to the present study. In this, and subsequent bandstructures, the position of the Fermi energy is indicated by a horizontal line.

FIGURE 4: Bandstructures and densities of states in fcc, bcc and hcp Cu using the frozen potential from self-consistent fcc Cu. Only the energy range above the d -bands is shown here. Symmetry points are those shown in figure 2.

FIGURE 5: Densities of states of fcc and bcc Cu superimposed to show the relationships between the important Van Hove singularities. Searching for those n at which fcc and bcc have the same Fermi energy leads to the two values indicated by vertical broken lines. At these points fcc and bcc respectively are at their most stable, the stable phase being that having the lowest density of states at the Fermi energy. The differences in densities of states can be attributed to interactions between the Fermi surface and the Brillouin zone planes marked in the figure.

FIGURE 6: bcc-fcc and hcp-fcc bandstructure energy differences using the rigid band structural energy difference theorem. Two frozen potentials have been used (see table 1): (a) self-consistent Cu potential, (b) self-consistent Zn potential. Crosses are the end-point total energy differences calculated selfconsistently within the local density approximation.

FIGURE 7: (a) The Cu-Zn phase diagram.³¹ (b) Free-energy-composition curves based on the data in figure 6(a). The curves are sheared using Zener's device: n_{bcc} is the valence electron concentration at which the two phases joined by the common tangent have the same energy, denoted u_{bcc} ; and u'_{bcc} is the slope of the bcc curve at $n = n_{\text{bcc}}$. A common tangent shows the positions of the zero-temperature phase boundaries deduced from the rigid band structural energy difference theorem. They cannot be interpreted in terms of the actual phase diagram since the α/γ equilibrium is not considered here; however, we show them to compare with Jones results.^{3,30}

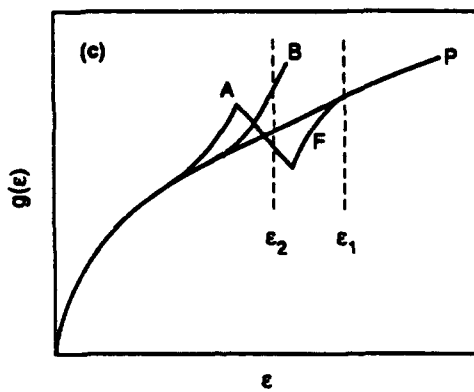
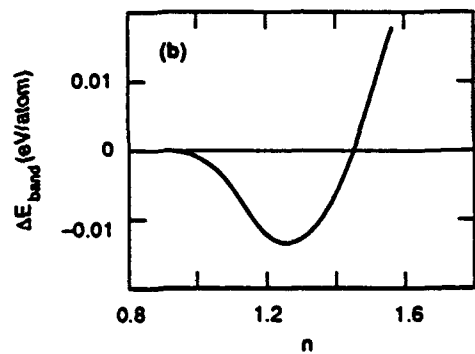
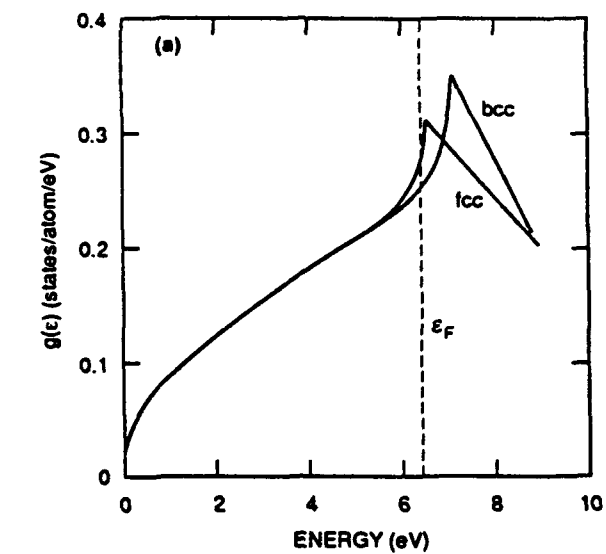
FIGURE 8: Analysis of the phase transitions in the rigid band structural energy difference theorem. (a) is the same as figure 7(b); (b) density of states as a function of valence electron concentration; (c) Fermi energy difference as a function of valence electron concentration. The units of energy we use in this paper are Rydbergs (1Ry = 13.6eV, and 1mRy/atom = 1.23kJ/g-atom).

FIGURE 9: Bandstructure and density of states using our parameters for Jones' model (table 1).

FIGURE 10: As figure 8, but with the band parameters of the Jones model

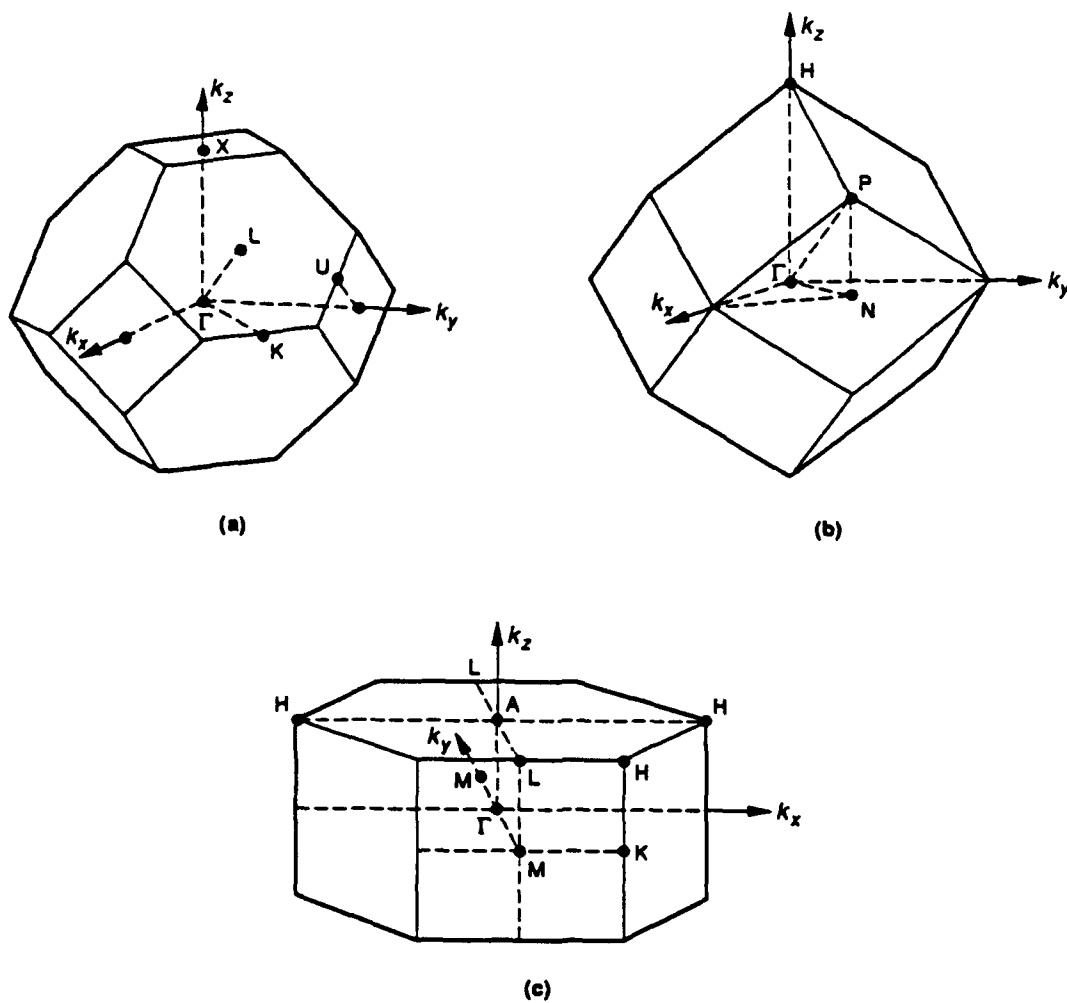
FIGURE 11: Bandstructure and density of states using our parameters for the Mott and Jones model (table 1).

FIGURE 12: As figure 8, but with the band parameters of the Mott and Jones model. Note the oscillations in the hcp density of states due to band crossings.



ps-120491-jd

Fig 1



ps-120291-jd

FIG. 2

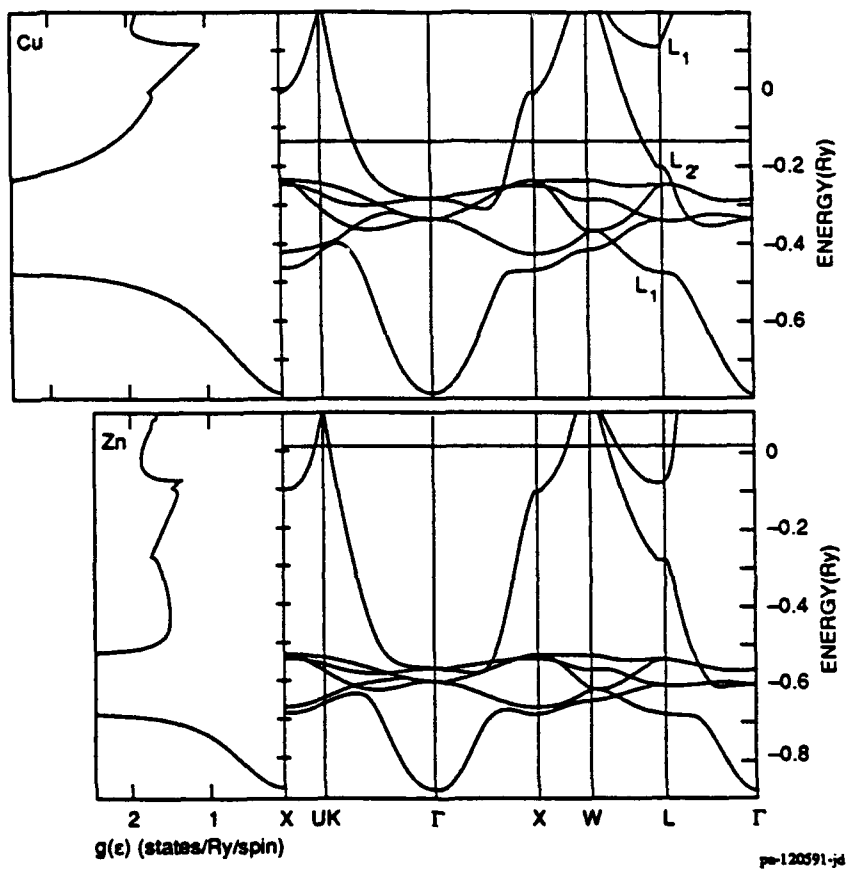
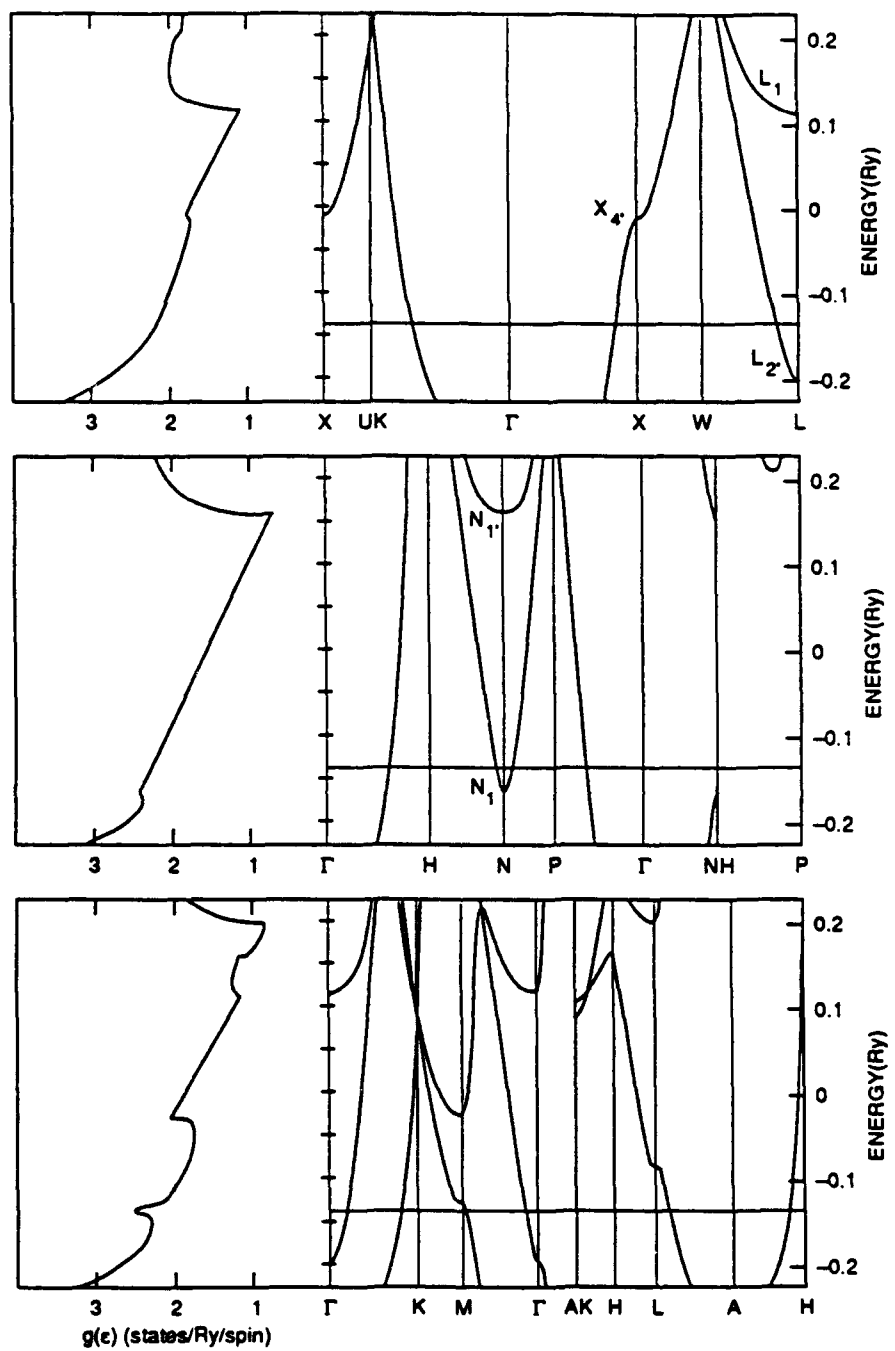


Fig. 3



ps-120891-jd

FIG. 4

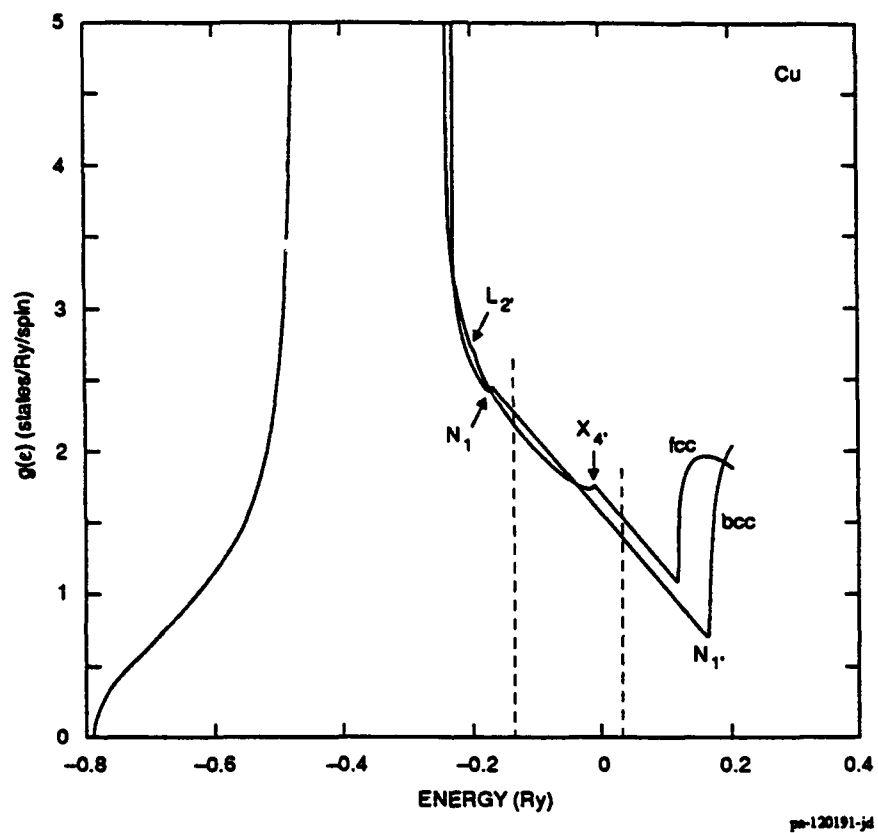


Fig. 5

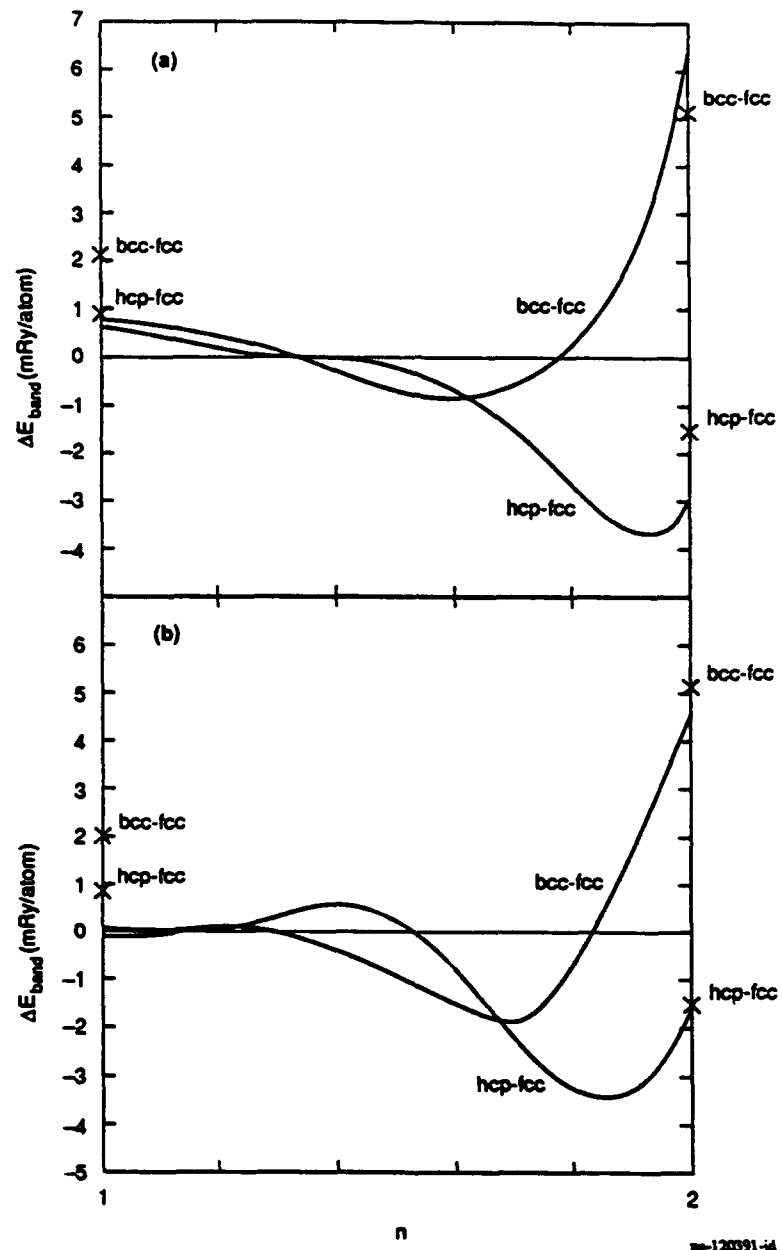
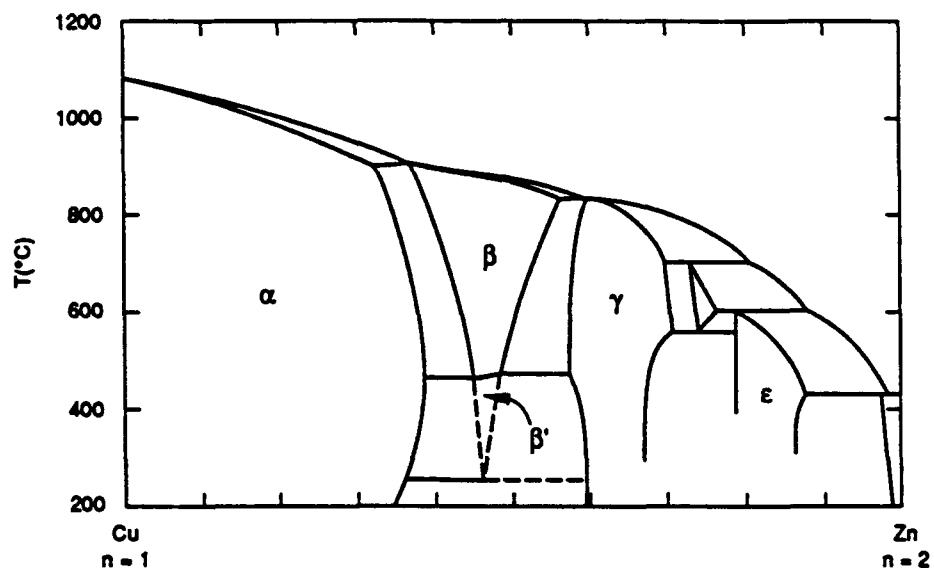
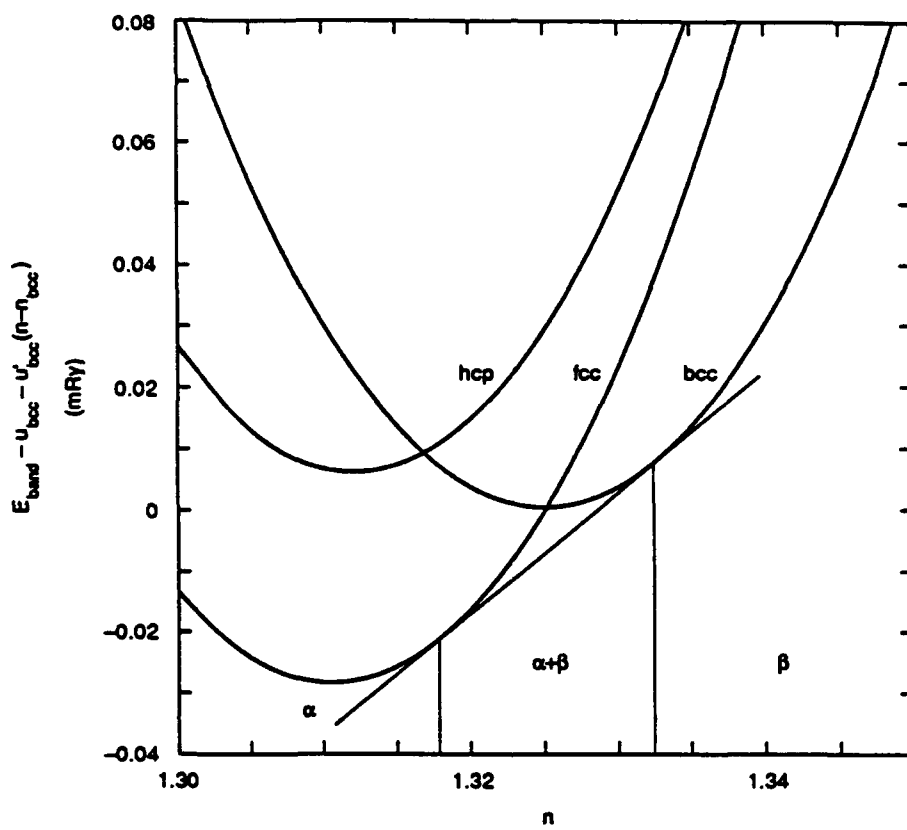


FIG. 6



pe-012992-jd

Fig 7a



pe-013093-jd

Fig 7b

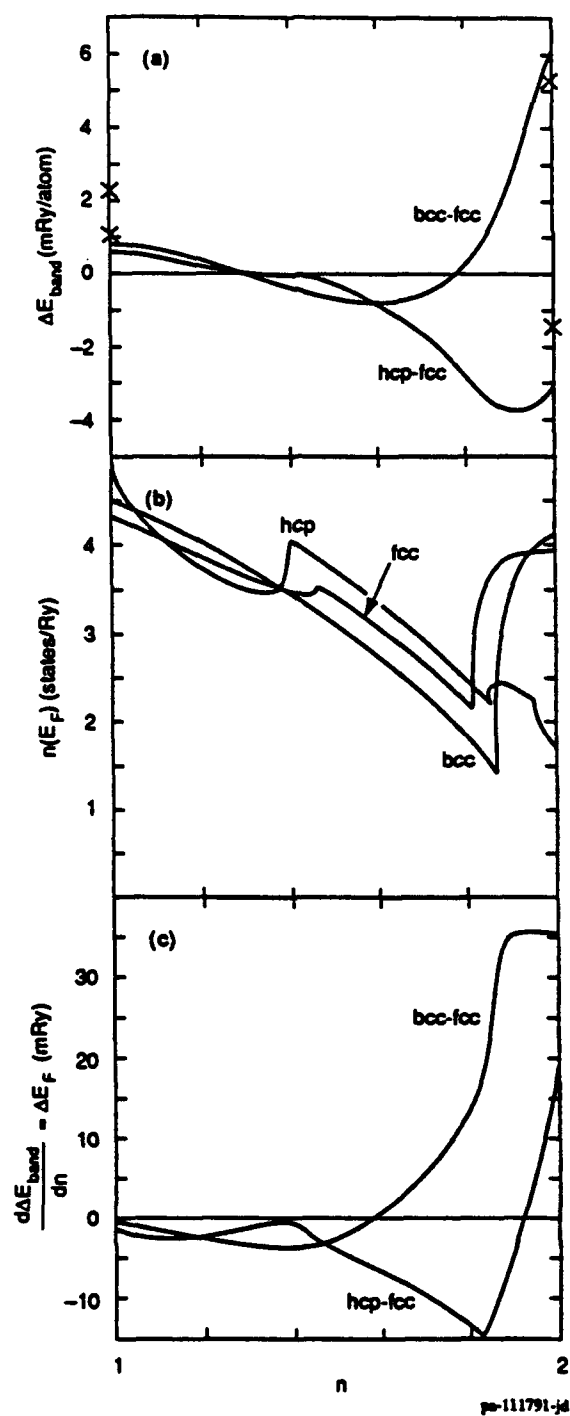


Fig 8

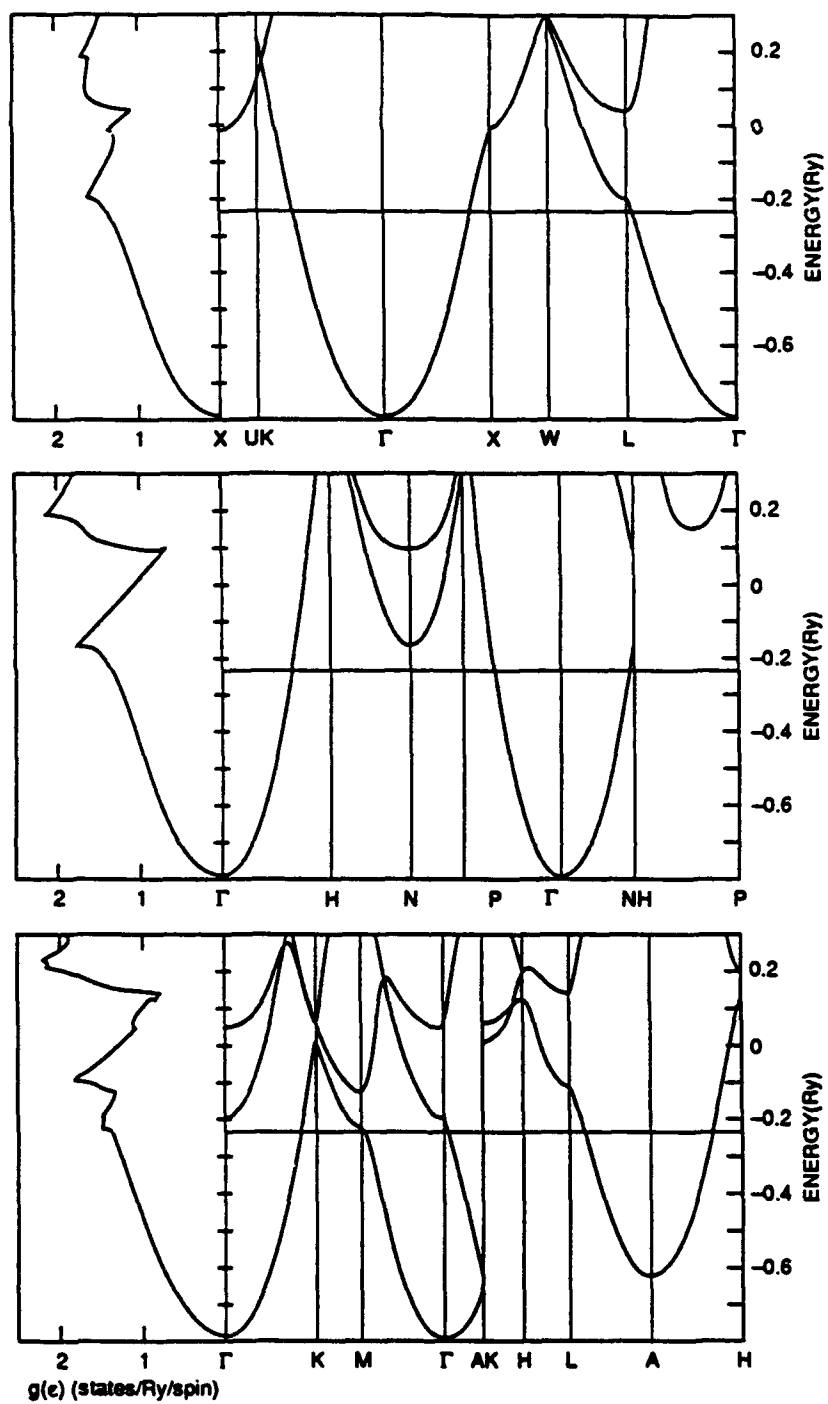


FIG. 9

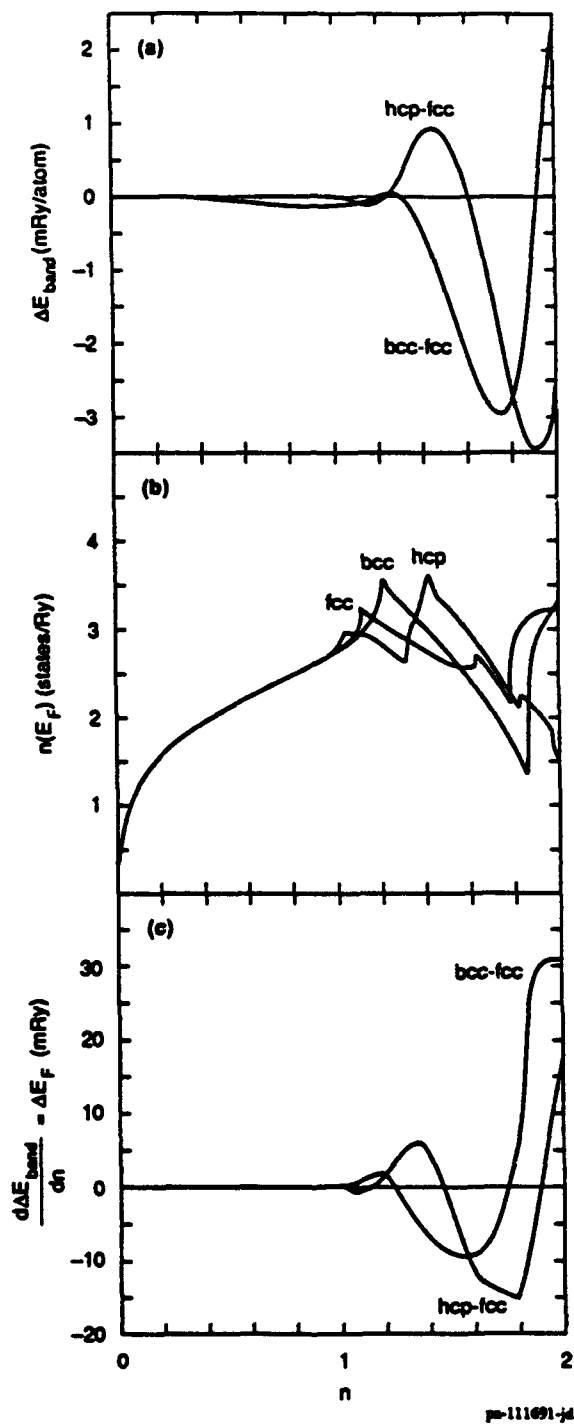
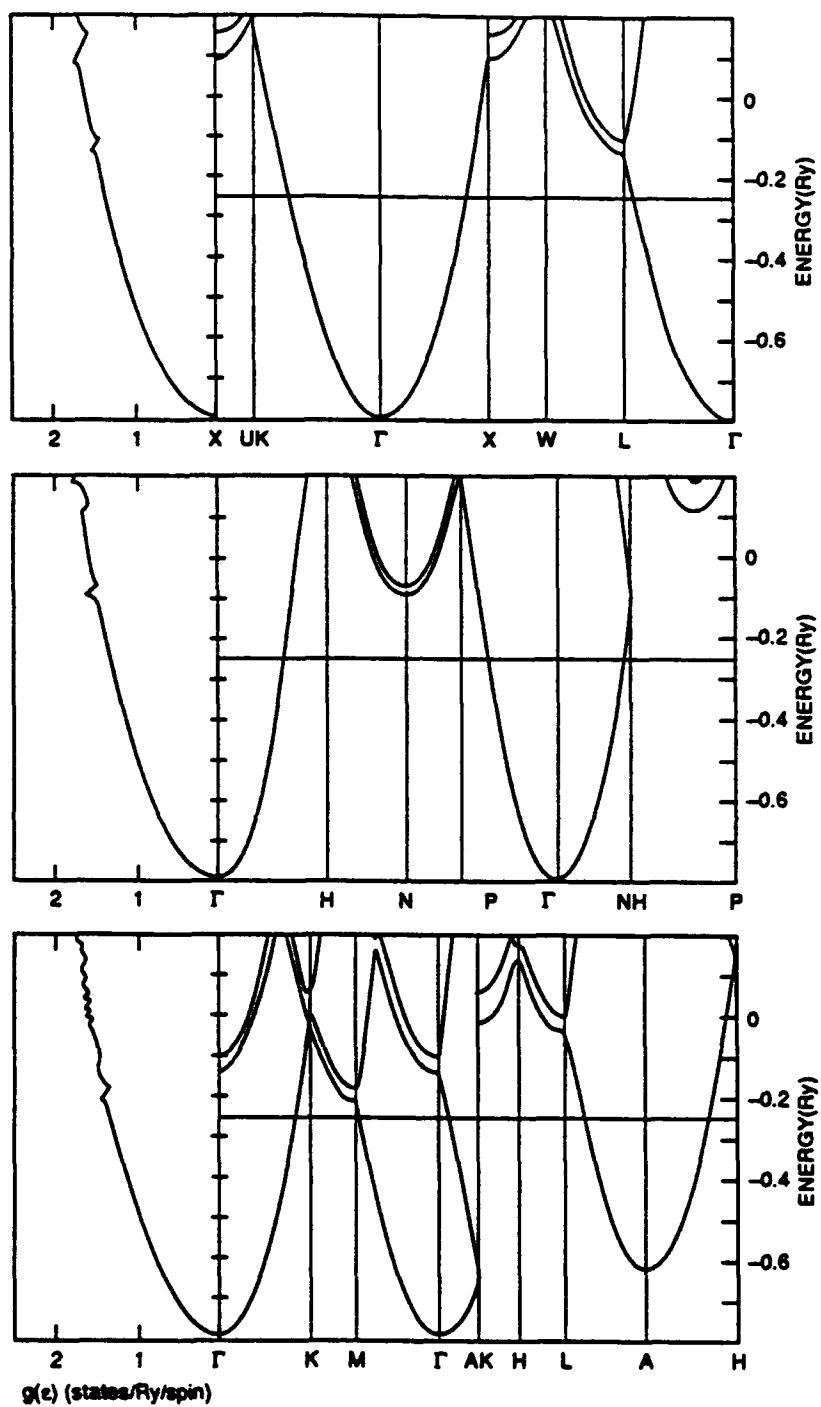


FIG 10



ps-12001-jd

FIG. 11

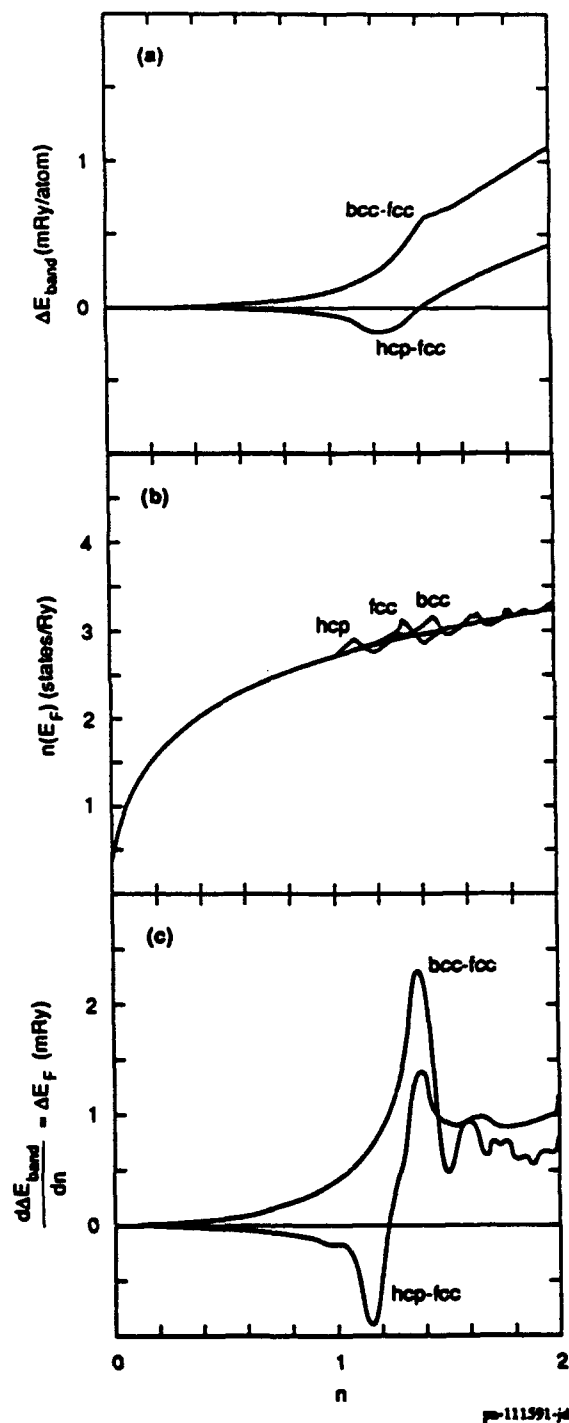


FIG. 12

Appendix I

FIRST PRINCIPLES DETERMINATION OF THE NI-AI PHASE DIAGRAM

First-principles determination of the Ni-Al phase diagram

A Pasturel†, C Colinet†, A T Paxton‡ and M van Schilfgaarde‡

† LTPCM, CNRS Unité de Recherche Associé 29, ENSEEG, BP 75,
38402 St Martin d'Hères Cédex, France

‡ SRI International, Menlo Park, CA 94025, USA

Received 16 July 1991, in final form 9 October 1991

Abstract. Phase stability in the Ni-Al binary system is investigated using linear muffin-tin orbitals total energy (LMTO) calculations. They provide total energies for the different existing compounds and, using Connolly-Williams inversion, the many-body interactions occurring in the FCC and BCC lattices. These interactions are used in conjunction with the Cluster Variation Method (CVM) to calculate the phase diagram. The computed phase diagram agrees very well with the experimental one.

1. Introduction

The modern theory of phase diagram calculations has been made possible by great advances in band-structure calculations, the theories of the configurational thermodynamics and phase transformations. Total energy calculations based on the local density approximation (LDA) are now sufficiently accurate to explain many properties of materials in terms of the underlying electronic structure [1]. An accurate calculation of the configurational free energy of the alloy is possible within various approximations such as mean field methods (Cluster Variation Method, CVM) or by numerical methods (Monte Carlo simulations) [2, 3]. In these models, it is assumed that the internal energy can be written as a sum of multi-site interactions which converge rapidly. The fact that these interactions can be derived from first-principle calculations establishes the basis of a comprehensive first-principles theory of cohesive, structural and thermodynamical properties of metals and alloys. Two extreme types of approach to the calculation of these interactions have been developed; the first one starts from the energy of the completely disordered solid solution calculated by the coherent potential approximation (CPA) [4]. The effective cluster interactions are calculated by the Embedded Cluster Method [5] or by the Generalized Perturbation Method (GPM) [6] using a perturbative treatment about the completely disordered state. In this case, the ordering energies can be written as an expression in terms of concentration-dependent n th-order effective cluster interactions. GPM can be developed with the first-principles multiple scattering formalism of the Korringa-Kohn-Rostoker coherent potential approximation [7]; however, to our knowledge, no phase diagrams have yet been provided by this approach although, in the framework of the tight-binding approximation, interesting results have been obtained [8-10].

The second approach is the so-called Connolly-Williams inversion [11] and the closely related ϵ - G approach [12]; here, the total energy is assumed to be written as a sum of configuration independent many-body interaction potentials multiplied by the multi-site correlation functions. The sum runs over all the cluster types and, in practice, it requires the existence of a maximum cluster beyond which the many body interactions are supposed to be negligible. This procedure has been used successfully by several authors [13-15] and we will adopt it here. The Ni-Al system presents both theoretical and technological interest, the aluminides having many desirable properties such as low density, high melting temperatures and high yield strength.

From a theoretical point of view, Sigli and Sanchez [16] have shown that a CVM treatment using effective pair interactions determined from available thermochemical data was able to give a good representation of the phase diagram. Several quantitative total energy calculations [17-20] have also been performed for ordered AlNi compounds using linearized methods [21]. All these calculations conclude that a strong hybridization between the p states of Al and the d states of Ni leads to the formation of bonding and antibonding states well separated by a pseudogap in the electronic density of states. The same trend has also been found using Tight Binding arguments [22]. A few calculations have treated more subtle effects such as the relative stabilities in Ni₃Al compounds [18]; van Schilfgaarde *et al* [23] have shown that non-self-consistent total energy calculations can be applied to this problem: they employ the linear muffin-tin method [21] using two different approaches. In the first approach, within the atomic sphere approximation, they employ an analogue of the local force theorem [24, 25], in which the total energy difference between two structures comprises the band-structure energy plus small Coulomb corrections. In the second approach, they make no shape approximations to the potential and employ a recasting of the local-density functional proposed by Harris and Foulkes [26]. All these results are in good agreement with experimental determinations of the energies of formation of Ni-Al compounds [27], which is the primary condition for the phase diagram determination. For the second approach, i.e. calculations of the cluster interactions, a series of results has been proposed by Carlsson, using Connolly-Williams inversion [28, 29] with a resummation scheme which leads to concentration dependent effective pair interaction (EPI). These so-obtained EPI are qualitatively comparable to the ones obtained from mean field methods based on multiple-scattering theory [30] and give very short-ranged interactions and strong ordering tendencies. A tight-binding approach based on the Cluster Bethe Lattice Method (CBLM) has also been used and gives EPI values very close to Carlsson's ones [22].

The purpose of this work is to present a complete calculation of the phase diagram of the binary NiAl system combining the tetrahedron approximation of the CVM and the linear muffin-tin orbital method (LMTO). This determination requires the knowledge of SRO in solid solutions based on BCC and FCC underlying lattices, the stability of the different compounds in the binary system and the liquid phase. The paper is organized as follows. In section 2, we present a brief review of the quantum and statistical mechanical approaches used in our calculations. In section 3, we present the results of our calculations and compare them with the experimental ones.

2. Free energy model

In order to compute a phase diagram, we need to know the total free energy of the binary alloy in a given phase, then its energy and also its entropy of formation.

For the alloy in phase α , the total free energy may be written as

$$F^\alpha = (1 - c)F_A^\alpha + cF_B^\alpha + E_f^\alpha - TS_f^\alpha \quad (1)$$

where c is the concentration of the B element, F_I^α , the free energy of pure element I in phase α , E_f^α and S_f^α are respectively the enthalpy and entropy of alloy formation.

The first simplifying assumption regarding the evaluation of this total free energy is to assume that the free energy for the pure elements can be written as

$$F_I^\alpha = E_I^\alpha - TS_I^\alpha \quad (2)$$

in which the cohesive energy E_I^α and entropy S_I^α are both temperature independent.

The second assumption assumes that S_f^α is a purely configurational term, which means that the remaining entropy depends only on the concentration via the first two terms on the right hand side of (1).

2.1. Energy of formation

It has been shown in the introduction that to perform phase diagram calculations, the internal energy of an A-B alloy is assumed to be described in terms of a rapidly convergent series of concentration-independent multi-site interactions. More precisely, we assume that the total energy of a particular configuration α is expressed by [31]

$$E_{\text{tot}}^\alpha(r) = \sum_{\gamma} V_{\gamma}(r) \xi_{\gamma}^\alpha \quad (3)$$

$V_{\gamma}(r)$ is the concentration-independent multi-site interaction associated with the multi-site correlation ξ_{γ}^α defined as [32]

$$\xi_{\gamma} = (1/N_{\gamma}) \sum_{\{n_j\}} \sigma_{n_1} \sigma_{n_2} \dots \sigma_{n_{\gamma}}$$

where $\sigma_n = 1 - 2p_n$, takes the values $+1$ or -1 depending on the occupancy of site n , N_{γ} is the total number of γ -type clusters, and the sum runs over all γ -type clusters that can be formed by combining sites on the entire crystal. The total energies and the multi-site interactions are generally lattice-parameter r (or volume) dependent.

Given (3), the interactions are determined by an inversion of the sum for a finite number of configurations defined by the existence of a maximum cluster γ_{max} beyond which the multi-site interactions vanish.

Hence, from a finite number of total energies for ordered structures and by truncating the summation in (3), a set of multi-site interactions is obtained from

$$\begin{aligned} V_{\gamma}(r) &= \sum_{\alpha} (\xi_{\gamma}^\alpha)^{-1} E_f^\alpha(r) & \Phi < \gamma < \gamma_{\text{max}} \\ V_{\gamma}(r) &= 0 & \gamma_{\text{max}} < \gamma < \infty \end{aligned} \quad (4)$$

Φ is the empty cluster.

Of course, one performs total energy calculations of as many ordered structures as there are unknown multi-site interactions required by the truncation but it is clear that, in this approach, this truncation leads to a non-uniqueness of the cluster interactions [32]. However, if the concentration-independent multi-site interactions decay rapidly, one can expect that this difficulty becomes minor and then in practice, the interactions can be uniquely computed.

In a series of papers, [28, 29, 33], Carlsson has computed cluster interactions for FCC and BCC alloys of aluminium with transition metals using Connolly-Williams inversion. For FCC calculations, he used A-FCC, B-FCC, A_3B-L1_2 , AB_3-L1_2 , and $AB-L1_0$ with ideal c/a ratio. Only first-nearest-neighbour pair interactions, triangle and tetrahedron cluster interactions are included. For BCC calculations, he takes A-BCC, B-BCC, A_3B-DO_3 , AB_3-DO_3 , AB-B2 and AB-B32. In this case, first- and second-nearest-neighbour pair interactions, triangle and tetrahedron cluster interactions are taken into account. He finds that this truncation is a reasonable approximation for these systems; for a Ni-Al system, the convergence of the cluster expansion appears to be very rapid, three- and four-atom terms having 15% and 5% of the magnitude of the pair terms [28]. This result is confirmed by the results for the effective pair interaction obtained via the Generalized Perturbation Method applied to the Korringa-Kohn-Rostoker multiple scattering formulation of the Coherent Potential Approximation (KKR-CPA) which finds roughly a factor of ten reduction at the second-neighbour distance [30].

Carlsson has also shown that it is possible to convert multi-site interactions into concentration dependent effective pair interactions [28]. They are obtained by making a truncated approximation in the higher-order correlation functions, using only the pair correlation function. Even if this development suffers from a loss of accuracy, the effective pair interactions present advantages in their ease of interpretability and their practical usefulness.

Then, using first-principles total energy calculations, we are able to compute the energies of formation of stoichiometric compounds occurring in the studied phase diagram and the multisite interactions allowing us to treat short-range order in the FCC- or BCC-based solid solutions or ordered superstructures presenting an extended concentration range.

2.2. Configurational entropy

As for the energies of formation, different configurational entropy approximations depending on the nature of the phase being considered are used. For the strictly stoichiometric compounds, the configurational entropy is taken equal to zero. For the solid solutions, or ordered phases presenting an extended concentration range, the configurational entropy is described by means of the CVM. The CVM entropy is found to be approximately given by a sum of the partial cluster entropies [34]. The maximum cluster used in our study is the tetrahedron containing first and second neighbours in the BCC lattice. In the tetrahedron approximation, the entropy of a BCC disordered system is given by [16]

$$S^{\text{BCC}} = -k_b \left(6 \sum_{ijkl} w_{ijkl} \ln w_{ijkl} - 12 \sum_{ijk} t_{ijk} \ln t_{ijk} + 3 \sum_{ij} y_{ij}^{(2)} \ln y_{ij}^{(2)} + 4 \sum_{ik} y_{ik}^{(1)} \ln y_{ik}^{(1)} + \sum_i x_i \ln x_i \right) \quad (5)$$

where w_{ijkl} , t_{ijk} , $y_{ij}^{(2)}$, $y_{ik}^{(1)}$ and x_i denote, respectively, the probability of finding tetrahedra, triangles, second-neighbour pairs, first-neighbour pairs and points in the configuration given by the subscripts (i equals A or B in a binary alloy).

For the disordered FCC structure we have

$$S^{\text{FCC}} = -k_b \left(2 \sum_{ijkl} w_{ijkl} \ln w_{ijkl} - 6 \sum_{ik} y_{ik}^{(1)} \ln y_{ik}^{(1)} + 5 \sum_i x_i \ln x_i \right). \quad (6)$$

The cluster probabilities are related by the following consistency relationships

$$t_{ijk} = \sum_l w_{ijkl} \quad (7a)$$

$$y_{ij}^{(2)} = \sum_{kl} w_{ijkl} \quad (7b)$$

$$y_{ik}^{(1)} = \sum_{jl} w_{ijkl} \quad (7c)$$

$$x_i = \sum_{jkl} w_{ijkl}. \quad (7d)$$

For the case of an ordered phase present in a range of concentration, long-range order is described in the usual manner by means of sublattices reflecting the symmetry of the ordered structure. A given cluster may now consist of points in the crystal belonging to different sublattices and their probabilities have to be distinguished accordingly (see [35] for more details).

2.3. Liquid alloys

Liquid alloys may also display SRO as has recently been shown for liquid $\text{Al}_{80}\text{Ni}_{20}$ alloy [36] and it is essential to consider SRO when determining the thermodynamic data. The best way to perform such calculations is to use a variational method with, as a reference system, a mixture of hard spheres which all have the same diameter but different charges and which interact through a screened Coulomb potential [37]. This reference system has been found, coupled with pseudo-potentials or TB-CBLM, to describe well the structural and thermodynamics manifestations of ordering in disordered alloys [38]. Very recently, it has been shown, using a tight-binding description of the total energy of the alloy [39] that, for the thermodynamic quantities, similar results are found for transition metal-based alloys if the liquid configurational free energy is approximated by the tetrahedron approximation of the CVM free energy in the disordered FCC structure. We have kept this approximation to describe the Ni-Al liquid phase in the present work.

2.4. Phase equilibrium

In order to determine the equilibrium phase diagram, it is more convenient to minimize the grand potential Ω given by

$$\Omega = F - \mu \xi_1 \quad (8)$$

where μ is the effective chemical potential. In the present work, the minimization of the grand potential is carried out with respect to a set of independent configurational variables and the variable r , since effective cluster interactions are also a function of the lattice parameter; in the use of the tetrahedron approximation, these configurational

variables are chosen to be the tetrahedron probabilities w_{ijkl} at constant temperature T and effective chemical potential μ , taking into account the normalization constraint

$$\sum_{ijkl} w_{ijkl} = 1. \quad (9)$$

This minimization is done using the Natural Iteration (NI) method developed by Kikuchi [40]. The NI equations used in the present model have been presented elsewhere [16] and will not be repeated here.

The equilibrium phase diagram between the two phases I and II is computed using the same scheme as proposed by Kikuchi and Murray [41]. For the same initial value of the effective chemical potential μ , the grand potentials of phases I (Ω_I) and II (Ω_{II}) are calculated using the procedure described above. If $\Omega_I = \Omega_{II}$, the equilibrium conditions are realized, but if not, the value of μ is modified until $\Omega_I = \Omega_{II}$.

3. Results for the Ni-Al system

The Ni-Al phase diagram displayed in figure 1 is characterized by a liquid phase, a FCC (A1) phase at both the Al and Ni rich ends, a non-congruently melting compound Al_3Ni and three intermediate phases with a variable range of solubilities Al_3Ni_2 , AlNi and AlNi_3 . Several attempts have been made to describe the NiAl phase diagram. As already mentioned, Sigli and Sanchez [16] have shown that the tetrahedron approximation of the CVM is an appropriate system to describe the phase diagram of this binary system.

In that case, EPI were determined from experimental thermodynamic data and available phase diagram information; these EPI are in good agreement with the ones determined by Carlsson [28] and Colinet *et al* [22] using quantum mechanics arguments.

3.1. Ground states at zero temperature

As a first step, we have calculated the energies of formation of the four intermediate phases observed in the equilibrium phase diagram. In the *strukturbericht* notation, these phases are called the orthorhombic DO_{20} (Al_3Ni) phase, the hexagonal D_{513} (Al_3Ni_2) phase, the cubic B2 (AlNi) phase and the cubic L1_2 (AlNi_3) phase. We have also selected two other FCC superstructures, L1_2 (Al_3Ni) phase and L1_0 (AlNi) phases and three other BCC superstructures, D0_3 (Al_3Ni), D0_3 (AlNi_3) and B32 (AlNi) phases. All these calculations allow us to check our prediction of the correct ground states and also to extract multi-site interactions to study the B2 (AlNi) and L1_2 (AlNi_3) phases, stable over an extended concentration range and described as SRO-phases. We mention that although the hexagonal D_{513} phase is experimentally reported to be stable over a small concentration range, it will be described as a stoichiometric compound.

To calculate the energies of the different phases, we have employed the all-electron total energy local density formalism as carried out with the linear muffin-tin orbital (LMTO) method [21]. The LMTO calculations have been done in the atomic sphere approximation, including combined corrections, as developed in the code of M van Schilfgaarde, A T Paxton and M Methfessel (unpublished result). We have used the parametrization of the von Barth Hedin exchange correlation energy density given by Moruzzi *et al* [42]. In our Brillouin zone integrals, we use a uniform mesh of sampling points with at least 16 divisions along each of the primitive vectors. With such a mesh, we obtain a convergence of the absolute value of the total energy of 0.1 mRy. The same

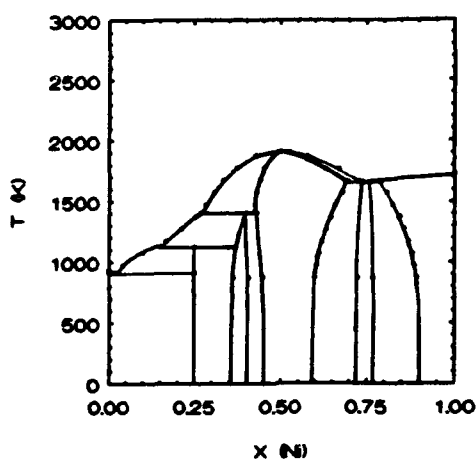


Figure 1. Experimental phase diagram of the Ni-Al system.

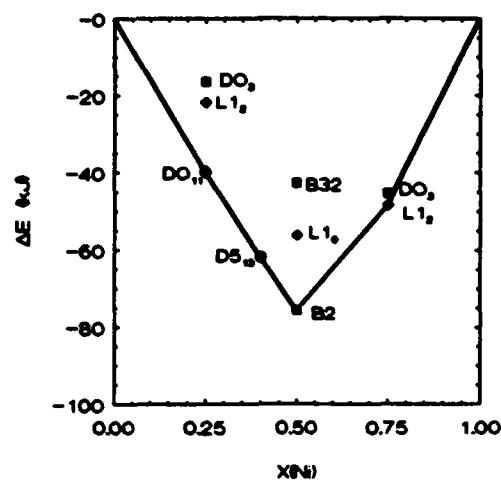


Figure 2. Formation energies as a function of composition for all the compounds studied.

Table 1. LMTO results for cohesive energies and equilibrium molar volumes of different structures in the Ni-Al system.

Structure	Composition	Cohesive energy (kJ mol ⁻¹)	Lattice parameter (au)
FCC	A ₁ Ni	-584.231	6.52
	L1 ₂ Ni ₃ Al	-586.732	6.675
	DO ₂₂ Ni ₃ Al	-584.074	6.685
	L1 ₀ NiAl	-548.698	6.828
	L1 ₂ NiAl ₃	-468.407	7.185
	DO ₂₂ NiAl ₃	-469.580	7.146
	A ₁ Al	-400.794	7.522
BCC	A ₂ Ni	-581.185	5.188
	DO ₃ Ni ₃ Al	-585.104	5.270
	B2 NiAl	-568.112	5.413
	B32 NiAl	-535.162	5.426
	DO ₃ NiAl ₃	-463.064	5.674
	A ₂ Al	-395.569	6.003
	D5 ₁₃ Ni ₂ Al ₃	-536.019	7.500
	DO ₂₀ NiAl ₃	-486.543	12.350

radius value was taken for the Wigner-Seitz spheres of all the elements and we include the spherical harmonic up to $l = 2$ (d orbitals) in constructing the basis functions.

For each structure, the total energies provided by the LMTO method are obtained for different values of the volume, the minimum of this curve determined the equilibrium total energy and the equilibrium volume. Moreover, the bulk modulus, which is related to the curvature of the total energy with volume is obtained using a fit based on Murnaghan's equation of state [43]. In table 1, are calculated equilibrium cohesive energies and molar volumes of all the considered phases. We recall that the cohesive

energy is defined as the difference of the total energy of that phase and the total energy of the constituent atoms at infinite interatomic distances. The ground states of both Ni and Al are predicted to be FCC. The energy difference between the FCC and BCC structures, defined as $\Delta E_{\text{FCC-BCC}} = E_{\text{coh}}^{\text{FCC}} - E_{\text{coh}}^{\text{BCC}}$, is 3.046 kJ mol⁻¹ and 5.224 kJ mol⁻¹ for Ni and Al respectively. We mention that the value of the structural energy difference for Al is very close to the one obtained by FLAPW calculations [15]. Figure 2 shows results for the formation energies as a function of composition for all the studied compounds. These formation energies are defined as

$$\Delta E = E_{\text{coh}} - x_{\text{Ni}} E_{\text{coh}}^{\text{FCC}}(\text{Ni}) - x_{\text{Al}} E_{\text{coh}}^{\text{FCC}}(\text{Al}). \quad (10)$$

At the equiatomic composition, the calculated formation energies of the competing phases, i.e. B2, B32 and L1₀ phases, strongly favour the B2 phase in complete agreement with experimental data. The computed value for ΔE of -75.6 kJ mol⁻¹ is more negative than the experimental value, -58.8 kJ mol⁻¹ [27]. However, our result is in complete agreement with other theoretical determinations [17] but this phase is known to present anti-sites and vacancies even at the equiatomic composition, factors which are not considered in these calculations. For 75% of Ni, the L1₂ phase is predicted to be more stable than the DO₃ phase, both being more stable with regard to a mixture of the Ni-FCC and NiAl-B2 phase. The computed value for ΔE of -48.36 kJ mol⁻¹ compares well with the experimental value, -41.0 kJ mol⁻¹. For 25% of Ni, the L1₂ phase is found to be yet more stable than the DO₃ phase, but its value is located well above the line connecting the formation energies of Al-FCC and NiAl-B2 phases. In fact for this composition, the most stable structure is the DO₂₀ orthorhombic phase, a result which is also obtained in our calculations. Its calculated formation energy, -39.89 kJ mol⁻¹ is in complete agreement with the experimental value, -37.7 kJ mol⁻¹.

At 40% of Ni for which the hexagonal D5₁₃ phase is experimentally reported to be stable over a small concentration range, there is no superstructure based on the BCC or FCC lattice. However, it is predicted to be more stable with regard to a mixture of NiAl₃-DO₂₀ and NiAl-B2 phases since the calculated formation energy, -61.85 kJ atom⁻¹, close to the experimental one, -56.5 kJ atom⁻¹, is just above the line connecting the formation energies of both DO₂₀ and B2 phases.

3.2. Cluster interactions and disordered alloys

As already mentioned, the concentration-independent multi-site interactions can be extracted from the total energy calculations using the Connolly-Williams inversion method.

In figures 3(a) and (b) are presented cluster interactions obtained for the FCC and BCC lattices as a function of the volume. These results reveal a very small difference between the FCC and BCC interaction parameters. V_0 and V_1 display a strong volume dependence. The shape of V_0 is similar to the one of the $E(V)$ function since V_0 is given by a sum of the energies of the different superstructures occurring on a given lattice. The shape of V_1 is essentially due to the fact that the two alloying elements present two different volumes. Let us mention that the strengths of V_3 and V_4 are much smaller than that of V_2 , indicating a reasonable degree of convergence already at the four atom supercell level. At this level of discussion we recall that in the tetrahedron approximation of the Connolly-Williams method for FCC-based structures, the L1₂ and DO₂₂ structures display the same correlations and then are degenerate. LMTO results provide that for $x_{\text{Ni}} = 0.25$, $E_{\text{L1}_2} - E_{\text{DO}_{22}} = +1.17$ kJ mol⁻¹ while for $x_{\text{Ni}} = 0.75$, $E_{\text{L1}_2} - E_{\text{DO}_{22}} =$

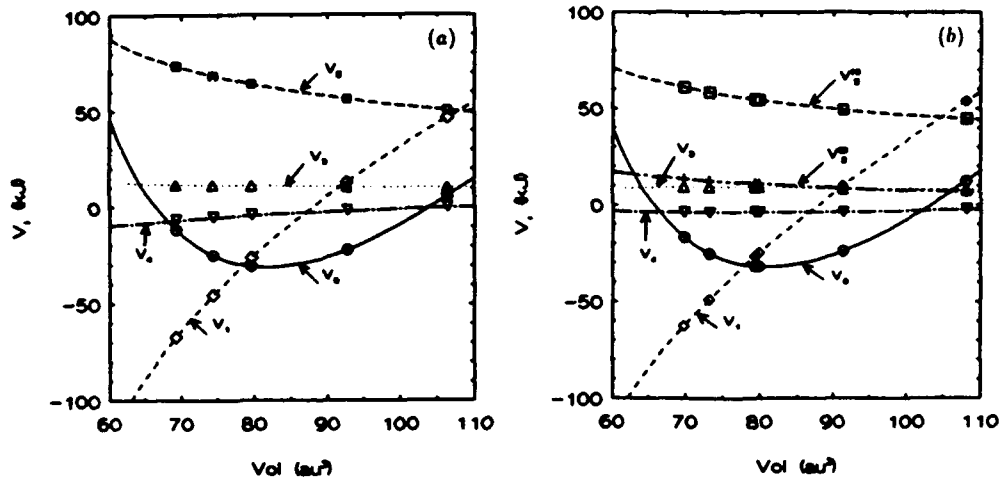


Figure 3. Cluster interactions as a function of the volume (a) FCC lattice; (b) BCC lattice.

$-2.66 \text{ kJ mol}^{-1}$. These two differences are small indicating also that the tetrahedron approximation is reasonable. It is interesting to compare the results obtained by LMT0 calculations to the ones provided by GPM-KKR-CPA [30] or TB-CBLM [22] approaches. This can be done by calculating the disordered energy and the effective pair interactions as a function of composition.

The disordered energy is obtained using the fact that the pair and higher order correlations are given as products of the point correlations for the totally disordered state. For the FCC and BCC-based structures, all atomic positions are equivalent and we have:

$$\xi_\gamma^{\text{dis}} = (\xi_1)^{n_\gamma} \quad (11)$$

where n_γ is the number of sites contained in the γ cluster. The energy of the disordered configuration is then given by

$$E_{\text{tot}}^{\text{dis}} = \sum_{\gamma} V_\gamma (\xi_1)^{n_\gamma} \quad (12)$$

Comparing cluster interactions from the CVM and effective pair interactions from GPM or CBLM can be done using a resummation of the higher order cluster interactions at fixed concentration, corresponding to a lowest order expansion of the total energy in powers of the short-range parameters [28].

With the set of clusters used in our calculations, we obtain for the FCC lattice

$$V_2^{\text{eff}} = \frac{1}{4}(V_2 + 3V_3\xi_1 + 6V_4\xi_1^2) \quad (13)$$

and for the BCC lattice:

$$\begin{aligned} V_2^{(1)\text{eff}} &= \frac{1}{4}(V_2^{(1)} + 2V_3\xi_1 + 4V_4\xi_1^2) \\ V_2^{(2)\text{eff}} &= \frac{1}{3}(V_2^{(2)} + V_3\xi_1 + 2V_4\xi_1^2) \end{aligned} \quad (14)$$

where the effective pair interactions V_2^{eff} are defined as

$$V_2^{\text{eff}} = \frac{1}{4}(V_{AA} + V_{BB} - 2V_{AB}) \quad (15)$$

V_{AA} , V_{BB} , V_{AB} are pair potentials evaluated at the separation between first-nearest

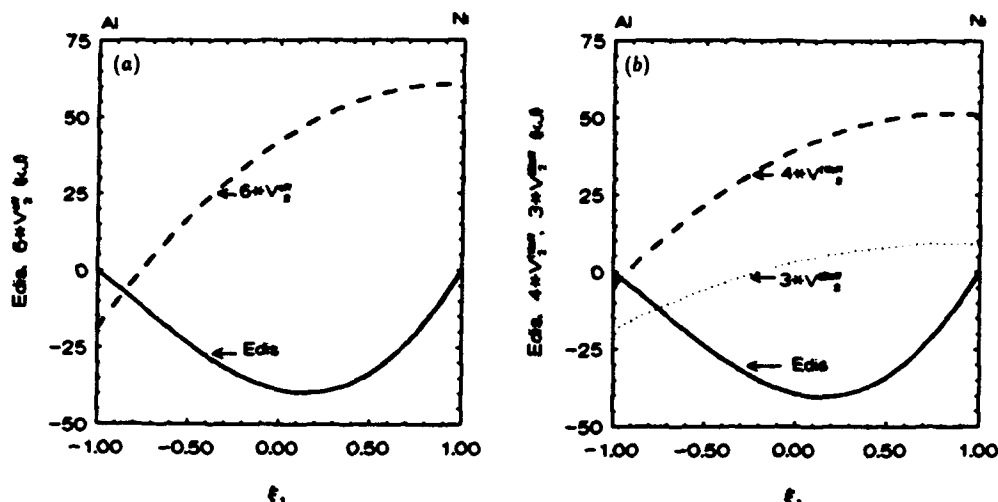


Figure 4. Concentration-dependent effective pair interactions and random energies (a) FCC lattice; (b) BCC lattice.

neighbours for the FCC lattice and first- and second-nearest neighbours for the BCC lattice.

As already mentioned by Carlsson, the EPI dependence as a function of the concentration is given by a higher order cluster than the pair here, triangle and tetrahedron. Two distinct approaches are used to obtain the concentration-dependent EPI: (i) the 'locally relaxed' treatment for which each cluster in the alloy is able to relax to its preferred lattice constant; (ii) the 'frozen lattice' treatment where each cluster is calculated at a fixed lattice constant. We use the 'locally relaxed' treatment of the lattice relaxation effects, which is considered to be more accurate than the 'frozen lattice' treatment [28]. In figure 4 are displayed values for the first-nearest-neighbour interactions on the FCC lattice as a function of ξ_1 (with ξ_1 equal to $x_{Ni}-x_{Al}$). Our results are of course very similar to Carlsson's ones since the augmented-spherical wave (ASW) and LMTO methods used to calculate total energies are essentially the same. Tight-binding based results present the same concentration dependence [22], and a large positive value of V_2^{eff} at the Ni rich end, consistent with the very strong ordering tendency in Ni_3Al which remains ordered up to its melting point; at the Al-rich end, the value of V_2^{eff} drops rapidly.

Our calculated EPI also display a semiquantitative agreement with tight-binding ones, $V_{FCC}^{(2)eff}$ being equal to 2.84 kJ, 7.02 kJ and 10.34 kJ for $x_{Ni} = 0.25, 0.5$ and 0.75 , respectively, while they are equal to 2.65 kJ, 4.58 kJ and 7.0 kJ in Colinet's approach. Another instructive comparison can be done with EPI determined by Turchi *et al* [30] in Ni-Al alloys around an equiatomic composition. These authors found $V_2^{(1)eff}$ (BCC) = 8.20 kJ for $x_{Ni} = 0.5$ compared to 9.97 kJ extracted from LMTO calculations or to 6.75 kJ obtained by Colinet *et al* [22].

In fact, Carlsson's resummation procedure corresponds to a high temperature expansion of the correlation functions and may be considered to be a good approximation to describe effective pair interactions in the liquid phase. For this kind of liquid alloy, experimental results based on neutron diffraction experiments [36, 44] have shown that

A-A and B-B distances are quite similar in the alloys but different from the nearest distances in pure liquids. Moreover it has been found that the local coordination is roughly equal to 11, very close to the FCC-coordination. Then, for each concentration, E_{dis} and V_{f}^{ff} used to describe the liquid part are obtained from (12), (13) at a fixed given volume. However, for each concentration a new volume is used. The concentration-dependent volume is taken to be similar to the one obtained for the FCC-based solid solution which displays a similar variation to the experimental one seen in the liquid phase [45]. Another argument to justify this approximation is that if the pair interactions of the FCC and BCC lattices come out similarly, then they are not sensitive to the underlying lattice (at least for the Ni-Al system).

In figure 5 are plotted these values of the V_{f}^{ff} as a function of the concentration. By comparison with the EPI obtained by the 'locally' relaxed treatment on the FCC and BCC lattice, we can see that these new values of V_{f}^{ff} display a smoother variation as a function of the composition.

The same conclusion is reached from the comparison of the random energy obtained in the locally relaxed treatment for the FCC and BCC lattices with the values obtained by this new treatment. From figures 4 and 5, we see that the three curves display negative values at all concentrations; the minimum is shifted towards the Ni rich part for the two curves obtained by the 'locally relaxed' treatment. The third curve displayed in figure 5, and used to describe the liquid phase, presents a more symmetrical shape.

3.3. Phase diagram calculations

Let us sum up the strategy of our calculations.

(i) The energies of the four compounds occurring in the phase diagram are obtained as a function of the volume. The minima of these curves are chosen to obtain the formation energies of these compounds. The entropies are considered to be equal to zero for the strictly stoichiometric compounds such as Al_3Ni , Al_3Ni_2 phases. For AlNi and AlNi_3 phases, the entropies of configuration will be given by the tetrahedron approximation of the CVM for ordered phases based on the FCC or BCC lattices.

(ii) The two solid solutions based on the FCC and BCC lattices are described as short-range-order solutions using the CVM treatment in its tetrahedron approximation. The cluster interactions are obtained as a function of the volume using the Connolly-Williams approach.

(iii) The difference in energy between FCC and BCC structures are taken from LMTO results.

(iv) The liquid phase is also described as a short-range-order phase using the tetrahedron approximation of the CVM free energy in the disordered FCC structure. The effective pair interactions are obtained from Carlsson's resummation procedure which corresponds to a high temperature expansion of the correlation functions. Although this model is developed with a solid-state approximation, it does a good job of describing the influence of chemical short-range order on the thermodynamic properties of the liquids [39, 46]. Of course, the continuous phase space of the nuclei's motion and structural information, such as pair correlation functions, are lost. However, the chemical ordering contribution to the entropy and the energy, which are quantities of central interest in our calculations, are only proportional to the concentration fluctuations. Using a statistical approach based on a 'discretized mesh' rather than the continuous

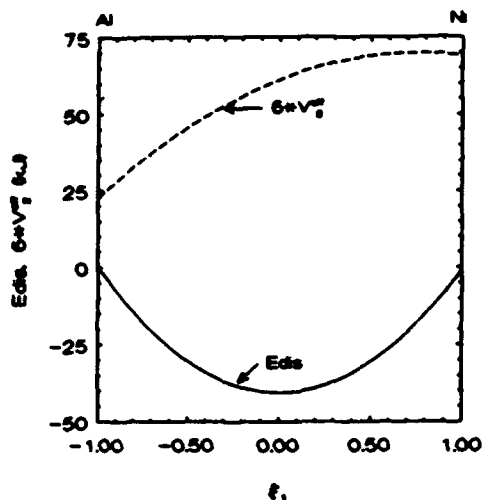


Figure 5. Concentration-dependent effective pair interactions and random energy for liquid phase.

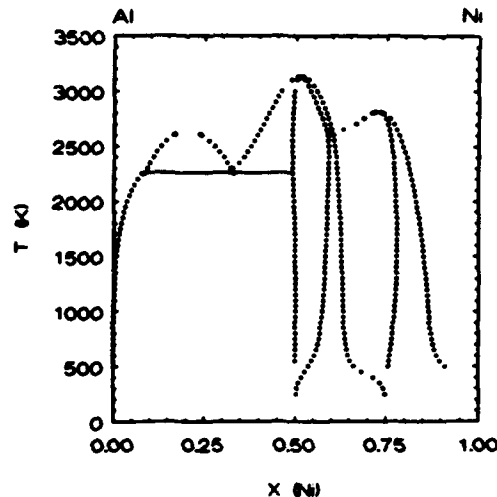


Figure 6. The Ni-Al phase diagram calculated with only FCC-based structures taken into account.

phase space of the nuclei's motion may be considered as correct to describe the concentration fluctuations.

However, there is still a 'missing link' in our approach which is the thermodynamic properties of the pure metals or, in other terms, the difference in free energy between the liquid and crystalline phases. Although the density functional theory has made significant progress in the modelling of liquids, application to the determination of the melting temperatures does not yet appear possible. Therefore, we have chosen to use thermodynamic compilations to obtain melting temperatures $T_{\text{FCC} \rightarrow \text{liq}}$ and the latent heat of melting $\Delta E_{\text{FCC} \rightarrow \text{liq}}$ for both Ni and Al elements [47]. In this case the free energy of these elements in their liquid state is given by

$$F_{\text{liq}}^{\text{el}}(T) = E_{\text{FCC}}^{\text{el}} + \Delta E_{\text{FCC} \rightarrow \text{liq}} - T(\Delta E_{\text{FCC} \rightarrow \text{liq}})/(T_{\text{FCC} \rightarrow \text{liq}}). \quad (16)$$

If only FCC-based equilibria are considered, the phase diagram of figure 6 is produced. Of course, only the Ni-rich portion can be directly compared with experimental results since for Ni concentration less than 0.75 structures other than FCC-based superstructures appear. The main features of the diagram are a miscibility gap (MG) with a maximum temperature of 2630 K and two ordered phases $L1_2$ and $L1_0$ with transition temperatures of 2820 and 3135 K respectively. As already mentioned by Carlsson and Sanchez [48], the MG is caused by an elastic instability and can be understood through the concentration dependence of the calculated energy of mixing for the completely random FCC solid solution. ΔE_{rand} is negative at all concentrations like the curves obtained with the 'locally relaxed' treatment (see figure 4) but here its curvature is negative for Ni concentrations less than 30%: this curvature is due to the peculiar volume dependence of V_2 , V_3 and V_4 , and another way to come to the same conclusion is to compute the phase diagram ignoring volume effects altogether. In that case the diagram displayed in figure 7 strongly resembles the previous diagram except for the MG on the Al-rich side.

When the families of both FCC and BCC-based free energy curves are combined with the stoichiometric Al_3Ni and Al_3Ni_2 compounds and the liquid phase, the phase diagram of figure 8 is obtained. All the main features of the experimental phase diagram (see

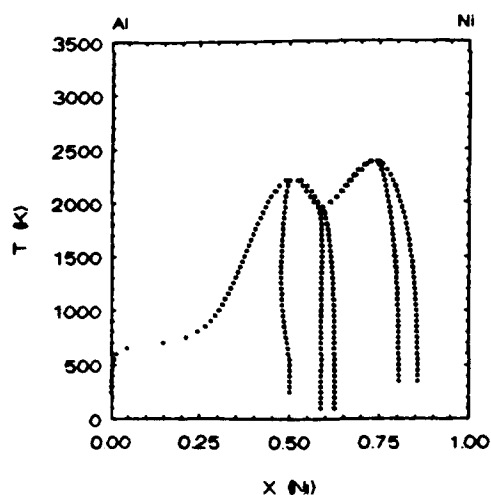


Figure 7. As figure 6 but now without including the volume dependence of the cluster interactions.

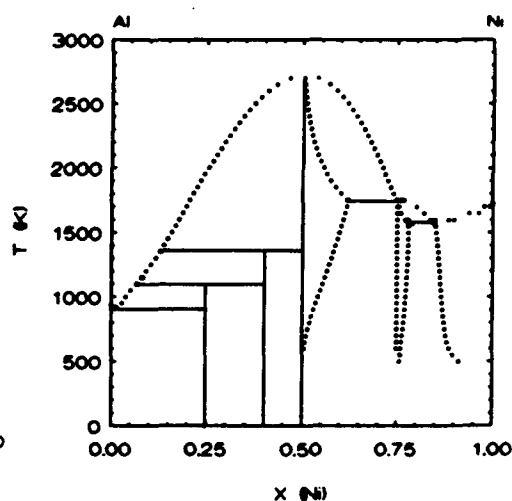


Figure 8. Calculated phase diagram of the Ni-Al system.

figure 1) have, at least qualitatively, been reproduced. The quantitative agreement is not as good.

(i) The B2 phase is found to melt congruently at $T = 2600$ K, before undergoing a disordering reaction. Its overestimate of the melting temperature seems to be attributable to the overestimate of the heat of formation of the B2 compound. However, in a more recent paper [49] Desai gives an experimental value of the heat of formation of the NiAl compound equal to $-71.65 \text{ kJ mol}^{-1}$ which is very close to our calculated value. He mentions the difference between the experimental values but no explanation is provided. Then the origin of the discrepancy of the calculated phase diagram with the experimental phase diagram in this region is not clear. Let us recall that this compound is known to present antisites and vacancies even at the equiatomic composition, factors which are not considered in these calculations.

(ii) The computed phase diagram exhibits two eutectics, one at the Al-rich side and another at the Ni-rich side, just like the experimental phase diagram. The calculated eutectic temperatures are in relatively good agreement with the experimental ones since they have been found equal to 900 K at the Al-rich side and 1680 K at the Ni-rich side (compared with 913 K and 1658 K respectively).

(iii) The two complex DO_{20} and DS_{13} structures are found to melt peritectically. For the Al_3Ni compound, the peritectic temperature is some 27 K lower than experiment indicates but for the Al_3Ni_2 compound it is 56 K higher than the experimental one.

Our calculations lead also to a peritectic decomposition of the Ni_3Al compound. More particularly, the eutectic between the solid phase A_1 , $\text{Ni}_3\text{Al}(\text{L}_{12})$ and the liquid one takes place on the right of the peritectic, which is in agreement with the usually adopted phases diagrams. However, let us mention that this part of the phase diagram is very sensitive to the values of the energies of the three phases which determine the equilibrium properties. From an experimental point of view, we point out that the phases boundaries of the peritectic decomposition of the Ni_3Al compound are still controversial.

4. Conclusion

It was shown that *ab initio* calculations of the Ni–Al phase diagram are, in many points, in agreement with experimental information. They have been obtained by combining total energy LMTO calculations with the tetrahedron CVM approximation. No relaxation and no vibrational effects have been taken into account. Only short-range order on FCC and BCC lattices have been introduced using CVM treatment in competition with the occurrence of complex phases like the DO₂₀ and D5₁₃ phases. The thermodynamic description of the liquid phase has been achieved by approximating this phase to a FCC-based disordered phase to minimize the number of parameters. The melting temperature and latent heat of the pure components are the only two parameters which have been introduced from outside the first-principles matrix. We believe that our results show that first-principles studies of phase equilibria may now be considered as feasible with a good degree of confidence.

Acknowledgment

ATP and MVS acknowledge financial support under AFSOR contract F49620-88-K-0009.

References

- [1] For a review, see Srivastava G P and Weaire D 1987 *Adv. Phys.* **36** 463
- [2] Kikuchi R 1951 *Phys. Rev.* **B 81** 988
- [3] Binder K 1986 *Monte Carlo Methods in Statistical Physics (Springer Series on Topics in Current Physics 7)* ed K Binder (Berlin: Springer)
- [4] Velicky B, Kirkpatrick S and Ehrenreich H 1968 *Phys. Rev.* **175** 747
- [5] Gonis A, Zhang X A, Freeman A J, Turchi P, Stocks G M and Nicholson D M 1987 *Phys. Rev.* **B 36** 4630
- [6] Ducastelle F and Gautier F 1976 *J. Phys. F: Met. Phys.* **6** 2039
- [7] Turchi P, Stocks G M, Butler W H, Nicholson D M and Gonis A 1988 *Phys. Rev.* **B 37** 5982
- [8] Sigli C, Kosugi M and Sanchez J M 1986 *Phys. Rev. Lett.* **57** 253
- [9] Sluiter M, Turchi P, Zchong Fu and De Fontaine D 1988 *Phys. Rev. Lett.* **60** 716
- [10] Le D H, Colinet C, Hicter P and Pasturel A 1991 *J. Phys.: Condens. Matter* **3** 7895
- [11] Connolly J W D and Williams A R 1983 *Phys. Rev.* **B 27** 5169
- [12] Ferreira L G, Mbaye A A and Zunger A 1987 *Phys. Rev.* **B 35** 6475
- [13] Zunger A, Wei S H, Mbaye A A and Ferreira L G 1988 *Acta Metall.* **36** 2239
- [14] Mohri T, Terakura K, Oguchi T and Watanabe K 1988 *Acta Metall.* **36** 547
- [15] Sluiter M, de Fontaine D, Guo X Q, Podloucky R and Freeman A J 1990 *Phys. Rev.* **B 42** 10460
- [16] Sigli C and Sanchez J M 1985 *Acta Metall.* **33** 1097
- [17] Hackenbracht D and Kübler J 1980 *J. Phys. F: Met. Phys.* **10** 427
- [18] Xu J H, Oguchi T and Freeman A J 1987 *Phys. Rev.* **B 36** 4186
- [19] Sarma D D, Speier W, Zeller R, Van Leuken E, de Groot R A and Fuggle J C 1989 *J. Phys.: Condens. Matter* **1** 9131
- [20] Hong T and Freeman A J 1991 *Phys. Rev.* **B 43** 6446
- [21] Andersen O K 1984 *NATO ASI on Electronic Structure of Complex Systems* ed P Phariseau and W T Temmerman (New York: Plenum) p 11
- [22] Colinet C, Bessoud A and Pasturel A 1989 *J. Phys.: Condens. Matter* **1** 5837
- [23] van Schilfhaarde M, Paxton A T, Pasturel A and Methfessel M 1991 *MRS Spring Meeting 1990* vol 186 ed G M Stocks, A F Giamei and D P Pope (Pittsburgh, PA: Materials Research Society) p 107
- [24] Pettifor D G 1978 *J. Chem. Phys.* **69** 2930
- [25] Andersen O K, Skriver H L, Nohl H and Johansson B 1980 *Pure Appl. Chem.* **52** 93
- [26] Harris J 1985 *Phys. Rev.* **B 31** 1770
- Foulkes W M C and Haydock R 1989 *Phys. Rev.* **B 39** 12520

- [27] Hultgren H, Desai P D, Hawkins D R, Gleiser M and Kelley K K 1973 *Selected Values of the Thermodynamic Properties of Binary Alloys* (Metals Park, OH: American Society for Metals)
- [28] Carlsson A E 1987 *Phys. Rev. B* **35** 4858
- [29] — 1987 High Temperature Ordered Intermetallic Alloys *Proc. Symp. Materials Research Society (Boston, 1986)* vol 81, ed N S Stoloff, C C Koch, C T Liu and O Izumi (Pittsburgh, PA: Materials Research Society) p 39
- [30] Turchi P, Sluiter M, Pinski F J and Johnson D D 1991 *MRS Spring Meeting 1990* vol 186 ed G M Stocks, A F Giamei and D P Pope (Pittsburgh, PA: Materials Research Society) p 59
- [31] Mohri T, Sanchez J M and de Fontaine D 1985 *Acta Metall.* **33** 1171
- [32] Sluiter M and Turchi P 1989 *Phys. Rev. B* **40** 11215
- [33] Carlsson A E 1989 *Phys. Rev. B* **40** 912
- [34] Kikuchi R and Sato H 1974 *Acta Metall.* **11** 1099
- [35] Kikuchi R and de Fontaine D 1978 *Nation. Bureau of Standards, Report* No SP 496, p 967
- [36] Maret M, Pomme T, Pasturel A and Chieux P 1990 *Phys. Rev. B* **42** 1598
- [37] Pasturel A, Hafner J and Hicter P 1985 *Phys. Rev. B* **32** 5009
- [38] Pasturel A and Hafner J 1986 *Phys. Rev. B* **34** 8357
- [39] Pasturel A 1989 unpublished results
- [40] Kikuchi R 1974 *J. Chem. Phys.* **60** 1071
- [41] Kikuchi R and Murray J 1985 *Calphad* **9** 311
- [42] Moruzzi V L, Janak J F and Williams A R 1978 *Calculated Electronic Properties of Metals* (New York: Pergamon) p 2
- [43] Murnaghan F D 1944 *Proc. Natl. Acad. Sci. USA* **30** 244
- [44] Maret M, Chieux P, Dubois J M and Pasturel A 1991 *J. Phys.: Condens. Matter* **3** 2801
- [45] Ayushina G D, Levin E S and Gole P V 1969 *Russian J. Phys. Chem.* **43** 11
- [46] Franz J R, Brouers F and Holzhey C 1980 *J. Phys. F: Met. Phys.* **10** 235
- [47] Dindsdale A T 1988 *National Physical Laboratory Report*
- [48] Carlsson A E and Sanchez J M 1988 *Solid State Commun.* **65** 527
- [49] Desai P D 1987 *J. Phys. Chem. Ref. Data* **16** 109

Appendix J

Al_3Ru —A DUCTILE TRIALUMINIDE?

Al₃Ru—A DUCTILE TRIALUMINIDE?

A.T.Paxton¹ and D.G.Pettifor²

¹SRI International, Menlo Park, California 94025

²Department of Mathematics, Imperial College, London SW7 2BZ

(Received October 30, 1991)

Introduction

Recent interest has focused on the transition metal trialuminides as candidate phases in the microstructure of high-temperature superalloys. They have excellent oxidation resistance, and the ordered, intermetallic compounds should have good high-temperature strength as do conventional cobalt and nickel based alloys. However, the compounds so far studied all suffer transgranular failure at room temperature, both those with cubic (L1₂) and non-cubic crystal structures (D0₂₂, D0₂₃ and D0₁₉). We need to understand and predict firstly the crystal structure adopted by transition metal trialuminides, and secondly their susceptibility to cleavage fracture. Cottrell (1) has shown that the latter can be rationalised within Pugh's analysis (2) and finds that a ductile crystal will have a ratio of shear to bulk moduli, $\mu/K < 0.4$; while a brittle crystal has $\mu/K > 0.5$. This has been confirmed for the L1₂ phases of Al₃Ti and Al₃Sc (3). The stable cubic transition metal trialuminides also tend to have a negative Cauchy pressure ($c_{12} - c_{44}$) which indicates the effect of angular bonding forces over and above the usual radial forces associated with the metallic bond (4). In this paper, we shall use the results of first-principles calculations to show that, on the basis of these criteria, Al₃Ru has a good likelihood of being the only ductile transition metal trialuminide.

The low temperature crystal structures of the transition metal trialuminides are shown in the lower part of fig. 1. The entry for Al₃Ru is taken from present predictions, otherwise the data come from refs. 5 and 6. We see that whereas group 3, 4 and 5 metals form stable close-packed trialuminides; of the others only Al₃Ru forms a close-packed phase. This is consistent with Carlsson's calculations of the three-body term, Φ_3 , in the Connolly-Williams expansion of the total energy (7). Carlsson found that Φ_3 oscillates across the 3d and 4d transition series, being negative in the neighbourhood of groups 3, 4, 7 and 8; and otherwise zero or positive. This corresponds with an observed skewing of the heats of formation away from regular solution behaviour (4, 8), stabilising the trialuminides of Sc, Ti, V and Ru; while as seen in fig. 1, elsewhere in the transition series close packed trialuminides are not observed.

We devote this paper to theoretical a study of Al₃Ru, in the light of the above remarks. Very little is known about the Al-Ru system. The phase diagram is only tentatively known (6). The structure of Al₃Ru is reported as D0₂₄ in Pearson's Handbook (5), with lattice parameters leading to an unphysically low atomic volume. Fortunately, the local density approximation (LDA) to density-functional theory is known to make accurate predictions of the crystal structures and elastic constants of crystals from first principles, and we exploit this using a basis of linear muffin-tin orbitals (LMTOs), in three distinct approaches (9, 10). (i) Using a full-potential, self-consistent method (11), we can obtain results corresponding to the LDA solution to Schrödinger's equation. (ii) In order to save computing time, we make the non self-consistent, Harris-Foulkes approximation when we are confident that the associated errors are acceptable. (iii) For predicting structural stability of close-packed phases, we use the atomic spheres approximation (ASA) to LMTO theory, using the structural energy difference theorem and rigid band approximation. This also allows us to construct a model to extract general principles arising from the electronic structure of transition metal trialuminides.

We divide the paper into two sections: crystal structure and elastic constant predictions. A final two sections contain our discussion and conclusions.

Crystal structures of the transition metal trialuminides

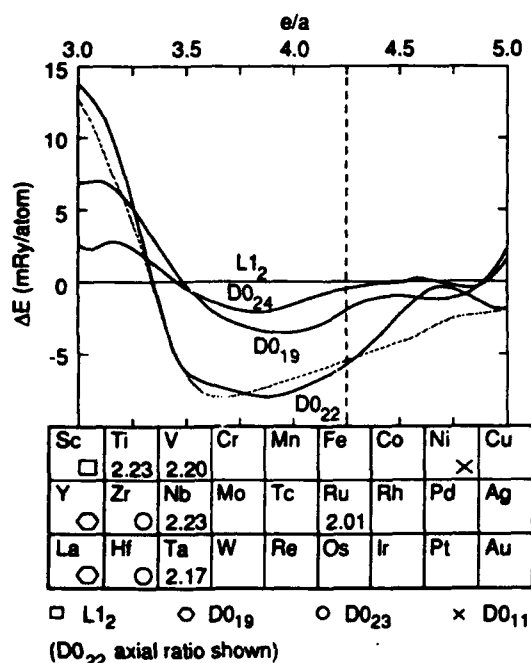


Fig.1 : The lower part shows those transition metals which form trialuminides with a symbol according to the legend given. The upper part shows calculated structural energy differences ($1 \text{ mRy/atom} = 1.29 \text{ kJ/g-atom}$). The solid curves show the energies of non-cubic trialuminides relative to $L1_2$ as a function of the electron concentration corresponding to the transition metals in the lower part. The rigid band approximation uses the DOS of Al_3Ru in the different structures, varying e/a while keeping the DOS unchanged, so that at the vertical broken line the calculations are not approximated. The dotted line shows the D0_{22} - $L1_2$ energy difference using the Al_3Ti DOS. Using this DOS we can predict the structure of Al_3Ru .

trial potential of the $L1_2$ phase, we have calculated the energies of D0_{22} , D0_{24} and D0_{19} (all with ideal axial ratios) relative to $L1_2$. These are shown at the intersection of the solid curves with the vertical broken line. In the spirit of the rigid band approximation, we use the Al_3Ru DOS to infer the energy ordering of all other transition metal trialuminides which results in the solid curves of energy difference ΔE against electron concentration e/a . Very much the same curves are obtained if we use the Al_3Ti DOS from fig. 4. For example, in fig. 1 the broken curve shows the D0_{22} - $L1_2$ energy difference obtained in this way. It is remarkable that using the electronic structure of either Al_3Ti or Al_3Ru , one can determine trends in the structural stability of all the other transition metal trialuminides. Our calculations shown in fig. 1 are for structures with ideal axial ratio. Therefore our model is not able to predict the structures of the group 3 and 4 trialuminides (apart from Al_3Sc). These have greater than 10% distortions from ideal axial ratio. As we see from fig. 1, it is this distortion that stabilises the observed phase;

[†]The cubic symmetry has split the five degenerate atomic d -states into sub-bands with e_g ($x^2 - y^2$, $3z^2 - 1$) and t_{2g} (xy , yz , zx) symmetry.

In order to calculate energy differences between competing close-packed phases, we use the structural energy difference theorem associated with the atomic spheres approximation (9). Beginning with a self-consistent calculation of the density of states (DOS) and atomic sphere potentials in the $L1_2$ structure, we transfer this *trial potential* into another structure and calculate its DOS. From the two densities of states, we calculate the sum of occupied energy levels and subtracting these, we obtain the structural energy difference correct to first order. It is instructive, first of all, to look at the DOS in Al_3Ru (fig. 2), since this demonstrates all the principal features of chemical bonding in the transition metal trialuminides: these are sketched in fig. 3. At the bottom of the band the electrons have predominantly free-electron-like Al- s character. The three peaks are typical of the transition metal trialuminides. The lower and upper peaks are bonding and antibonding hybrid Al- p -Ru- $d(t_{2g})$ states.[†] The central peak is predominantly nonbonding Ru- $d(e_g)$ (the peak is positioned just 0.05 Ry above the centre of the d -band indicating very weak antibonding character). The lower peak also has a bonding Al- p -Ru- $d(e_g)$ contribution. In the transition metal trialuminides, the bonding levels are full and the antibonding levels empty leading to strong, directional bonding between the transition metal and Al atoms. There is also metallic bonding mitigated by the free-electron-like Al and transition metal s -electrons. In fig. 4, we show the DOS for $L1_2$ Al_3Ti . Essentially the same features are present: the d -band is shifted to higher energy (the d -electrons become more tightly bound to the nucleus across the transition series) and the nonbonding states are empty since Ti has 4 fewer d -electrons than Ru. This leads one to the development of the rigid band model (12). Here we suppose that we may describe the electronic structure of the transition metal trialuminides using a rigid DOS, say, that of Al_3Ru ; and model all the transition metal trialuminides using this DOS, varying the Fermi energy to account for the total number of electrons.

Now consider the upper part of fig. 1. Using the DOS of Al_3Ru in four close-packed structures obtained from the

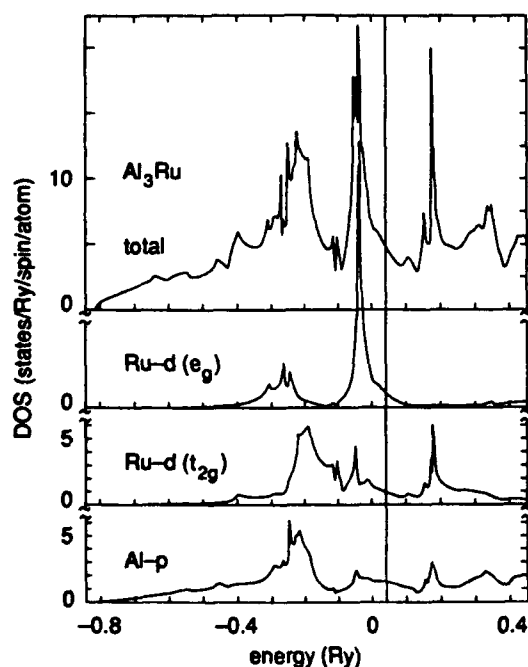


Fig. 2 : LMTO-ASA density of states of $L1_2$ Al_3Ru . A vertical line shows the position of the Fermi energy which separates occupied from unoccupied energy levels. The lower panels show the DOS decomposed into orbitals to demonstrate the principal features of the chemical bonding (see the text, and fig. 3).

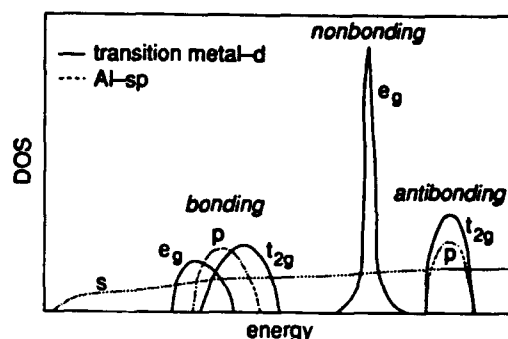


Fig. 3 : A schematic drawing of the important features of the DOSs shown in fig. 2. See the text for details.

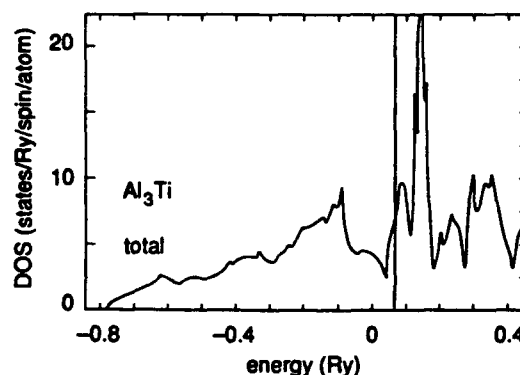


Fig. 4 : Density of states in $L1_2$ Al_3Ti . The Fermi energy is indicated as in fig. 2.

this is already known in the case of Al_3Ti from total energy calculations (4, 9). Very importantly, we see that while the group 4 trialuminides are on the borderline between $L1_2$ and $D0_{22}$ and can be made cubic by alloying (13), the group 5 compounds cannot (14).

Our calculations show that Al_3Ru has the $D0_{22}$ structure, not $D0_{24}$ as reported (5).[†] Furthermore, we see a very wide region of $D0_{22}$ stability around this electron concentration, so we regard it as highly improbable that Al_3Ru can be made cubic by alloying.

Elastic constant calculations

We need to make the most accurate total energy calculations to calculate elastic constants (11). Not only the band terms, as in the structural energy difference theorem, but also electrostatic and exchange-correlation terms must be included since these do not cancel in energy differences to second order. For shear modulus calculations we make a series of calculations of the total energy at a fixed amount of a pure strain. We vary the strain in eleven evenly-spaced increments in the range ± 0.05 to obtain the total energy as a function of strain. The second order terms are extracted by least-squares fitting to polynomials. Very stringent convergence is required since the energy varies by typically only a few mRy over the whole range of distortions, and great care is needed if one is to obtain a smooth fit to the data. Typically, our self-consistent energies are converged to better than 0.01 mRy,

[†]We should emphasise that our model is strictly concerned with trialuminides. We cannot make any statements concerning phases such as $Al_{13}Fe_4$ which has nearly the same stoichiometry and is sometimes referred to as Al_3Fe . A similar phase in the Al-Ru system may also exist.

and we employ a modified tetrahedron method of Brillouin zone integration (15) using a uniform mesh of at least 18^3 k-points in the whole zone. The bulk modulus is calculated from nine separate dilatations in the range ± 0.2 ; since the energy changes by much larger amounts convergence is less critical here. We divide the presentation of our results into those for $L1_2$ and $D0_{22}$ structures.

$L1_2$ structure

Although, as we have shown, Al_3Ru is never likely to exist in the $L1_2$ structure, it is instructive at first to examine its three elastic constants in that phase. Conveniently, these are the bulk modulus $K = \frac{1}{3}(c_{11} + 2c_{12})$, and the two shear constants $C = c_{44}$ and $C' = \frac{1}{2}(c_{11} - c_{12})$. In table 1, we show these, the Cauchy pressure and the ratio μ_2/K where $\mu_2 = (3c_{44}(c_{11} - c_{12})) / (4c_{44} + c_{11} - c_{12})$ is the reciprocal of the elastic compliance of a cubic crystal orientated for $[11\bar{2}]$ or $[\bar{1}10]$ slip on (111) planes (16). We also show the same quantities we have calculated for Al_3Ti . We see that Al_3Ru differs significantly from Al_3Ti , in that it satisfies the criteria for ductility discussed in the introduction, namely a positive Cauchy pressure and $\mu/K < 0.4$.

$D0_{22}$ structure

In table 1, we have also shown in parentheses the elastic constants for Al_3Ru calculated non self-consistently using the Harris-Foulkes approximation (9, 10). There is clearly negligible error involved, while the saving in computer time is considerable. For calculations in the $D0_{22}$ phase of Al_3Ru , we continue to use the non self-consistent approach. We show our calculated lattice constants in table 2; and in table 3, calculated elastic constants of Al_3Ru , which we compare with the measured elastic constants of Al_3Ti (17).

In table 3, the last four columns show firstly the two Cauchy discrepancies, and secondly the μ/K ratios for the known slip systems in $D0_{22}$ alloys. $\mu_1 = c_{44}$ is the reciprocal compliance in $\frac{1}{2}a[110](001)$ slip; and μ_2 , with the same expression as given above for cubic crystals, is for $a[110]$ slip on the close-packed $(1\bar{1}2)$ planes. The former slip system is operative in $D0_{22}$ Al_3Ti , while the latter is observed in Al_3V (18).

TABLE 3

Elastic Constants of $D0_{22}$ Al_3Ru and Al_3Ti

	K	c_{11}	c_{33}	c_{12}	c_{13}	c_{44}	c_{66}	$c_{13} - c_{44}$	$c_{12} - c_{66}$	μ_1/K	μ_2/K
	GPa										
$Al_3Ru^{(a)}$	164	236	259	143	115	106	89	9	54	0.65	0.35
$Al_3Ti^{(b)}$	106	218	218	58	46	92	117	-46	-59	0.87	0.79

(a) Calculated in the present work
(b) Measured in ref. (17)

(a) Calculated in the present work

(b) Measured in ref. (17)

TABLE 1

Calculated Elastic Constants of $L1_2$ Al_3Ru and Al_3Ti

	C	C'	K	$c_{12} - c_{44}$	μ_2/K
	GPa				
Al_3Ru	103	38	160	31	0.30
	(101)	(35)	(160)		
Al_3Ti	87	68	117	-15	0.63

TABLE 2

Lattice Parameters of $D0_{22}$ Al_3Ru and Al_3Ti

	a	c/a
	(nm)	
$Al_3Ru^{(a)}$	0.381	2.01
$Al_3Ti^{(b)}$	0.378	2.23

(a) Calculated in the present work

(b) Calculated in ref. (9)

Discussion

We need to answer the question, do our calculations indicate that Al_3Ru deserves some experimental attention? It is usually supposed that only cubic intermetallics will have sufficient room temperature ductility to be useful, and we have seen that it is most unlikely that an $L1_2$ form of Al_3Ru can be stabilised by alloying. Two points are at issue here, lack of ductility due to violation of Von Mises criterion; and intrinsic lack of cleavage strength. The slip system $a[110](1\bar{1}2)$ favoured by Al_3V (and, we expect, also Al_3Ru) has only four, rather than the five independent modes necessary for general ductility; this may however be sufficient, particularly if augmented with basal slip for which the μ/K ratio is more favourable in Al_3Ru than Al_3Ti , which is seen to have a very unfavourable μ/K ratio on its chosen slip system. This would explain the brittleness of Al_3Ti , although we don't yet understand what determines the choice of slip system in the transition metal trialuminides. Our calculations show significant differences between the elastic properties of Al_3Ru and Al_3Ti . The negative Cauchy discrepancies in both $L1_2$

and $D0_{22}$ Al_3Ti indicate large angular forces and low bond mobility. In Al_3Ru , although angular forces are still expected to play an important rôle (8), the elastic constants in $L1_2$ are rather similar to those of the ductile crystal α' - Ni_3Al . It is important, also, to note that whereas all the known $D0_{22}$ and $D0_{23}$ transition metal trialuminides have axial ratios 10–15% larger than ideal, Al_3Ru is predicted to have an axial ratio very nearly ideal, so that the (112) planes are genuinely close-packed. These observations lead us to expect that a ductile $D0_{22}$ alloy of Al_3Ru , possibly with ternary additions, can be found.

An important influence will be the anti-phase boundary (APB) energy on (112) planes. If this is low in Al_3Ru compared with Al_3V , then partial dislocations with half the Burgers vector of those observed in Al_3V will be responsible for the deformation, with a consequent enhancement of ductility. In future work, we will concentrate on a comparison of elastic constants and APB energies in Al_3Ru and Al_3V . This will give a better indication of the promise of Al_3Ru as a ductile trialuminide, and also assist our understanding of the choice of slip and twinning systems in the transition metal trialuminides.

Conclusions

By taking a flexible approach to density-functional theory in the local density approximation, we can make models to infer general trends, and highly accurate calculations to predict specific physical properties of inter-metallic compounds.

We have applied this to the transition metal trialuminides to get a picture of their chemical bonding and to see general trends in their phase stability, and also to take a detailed look at the lattice and elastic constants of Al_3Ru . We have demonstrated the remarkable fact that, using the structural energy difference theorem and rigid band model, one can predict the energy ordering in Al_3Ru using the electronic structure of Al_3Ti , and *vice versa*. The results of the rigid band calculations are also consistent with the trends in the phase stability of the transition metal trialuminides, given that axial ratio distortions in the non cubic phases are unaccounted for in our model. Accurate, full-potential calculations indicate distinct differences between Al_3Ru and Al_3Ti . We expect that Al_3Ru will deform similarly to Al_3V , but may well have enhanced ductility since it has ideal axial ratio. Differences in APB energy on the close-packed planes will be investigated in future work to confirm or deny this. Unlike other transition metal trialuminides, Al_3Ru shows some promise of room temperature ductility, even though it, and its ternary alloys, are unlikely to have a structure other than $D0_{22}$.

Acknowledgments

We thank Michael Methfessel for the use of his full-potential LMTO programs, and Julie Dubbs for preparing the figures. ATP gratefully acknowledges support under AFOSR contract F49620-88-K-0009.

References

1. A. H. Cottrell, *Proc. European Conf. Advanced Materials and Processes*, Cambridge, 22 July 1991
2. S. F. Pugh, *Phil. Mag.*, **45**, 823 (1954)
3. C. L. Fu, *J. Mater. Res.*, **5**, 971 (1990)
4. D. G. Pettifor, *NATO ASI on Ordered Intermetallics—Physical Metallurgy and Mechanical Behaviour*, Irsee, Germany, June 1991, C. T. Liu and R.W. Cahn (eds.) Kluwer Academic Publishers (to be published)
5. P. Villars and L. D. Calvert, (eds.), *Pearson's handbook of crystallographic data for intermetallic phases*, (ASM, Ohio, 1985)
6. T. B. Massalski *et al.*, (eds.), *Binary alloy phase diagrams*, (ASM, Ohio, 1986)
7. A. E. Carlsson, *Phys. Rev. B*, **40**, 912 (1989)
8. D. G. Pettifor and M. Aoki, *Phil. Trans. Roy. Soc. A*, **334**, 439 (1991)
9. M. van Schilfgaarde *et al.*, *Mat. Res. Soc. Symp. Proc. Vol. 186* (1990) and references therein
10. A. T. Paxton, M. Methfessel and H. M. Polatoglou, *Phys. Rev. B*, **41**, 8127 (1990)
11. M. Methfessel, *Phys. Rev. B*, **38**, 1537 (1988)
12. T. B. Massalski and U. Mizutani, *Prog. Mater. Sci.*, **22**, 151 (1978)
13. J. H. Schneibel and W. D. Porter, *Mat. Res. Soc. Symp. Proc. Vol. 133*, p. 335 (1989); W. D. Porter, *et al.*, *ibid.*, p. 657
14. P. R. Subramanian, *et al.*, *Mat. Res. Soc. Symp. Proc. Vol. 133*, p. 51 (1989)
15. P. E. Blöchl, Ph.D. Thesis, University of Stuttgart, 1989
16. A. Kelly and N. H. Macmillan, *Strong Solids*, 3rd. ed., pp 28, 389, Clarendon Press, Oxford (1986)
17. M. Nakamura and K. Kimura, *J. Mat. Sci.*, **26**, 2208 (1991)
18. R. Wheeler, V. K. Vasudevan and H. L. Fraser, *Phil. Mag. A*, **62**, 143 (1990)